

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

**Instituto de Ciencias Matemáticas
Ingeniería en Estadística Informática**

“Procedimientos estadísticos multivariados para el análisis de datos biológicos recopilados en el tiempo; sobre el desarrollo de una enfermedad foliar en banano y su relación con parámetros de clima”.

TESIS DE GRADO

Previo a la Obtención del Título de:

INGENIERO EN ESTADÍSTICA INFORMÁTICA

Presentado por:

Freddy Samuel Navarro Tobar

GUAYAQUIL – ECUADOR

Año 2008

AGRADECIMIENTO

A Dios por estar siempre conmigo, a mis padres por sus esfuerzos, confianza y amistad, a mi esposa e hijo por su entrega, a mis hermanas que siempre me apoyaron y a mi Director Ing. Omar Ruiz que me supo orientar con sus conocimientos y experiencias.

DEDICATORIA

A mis Padres,

A mi Esposa e Hijo,

A mis Hermanas.

TRIBUNAL DE GRADUACIÓN

ING. LUIS RODRÍGUEZ
PRESIDENTE DEL TRIBUNAL

ING. OMAR RUIZ
DIRECTOR DE TESIS

ING. FÉLIX RAMÍREZ
VOCAL

ING. GUILLERMO BAQUERIZO
VOCAL

DECLARACIÓN EXPRESA

“La responsabilidad del contenido de esta tesis de grado, me corresponde exclusivamente; y el patrimonio intelectual de la misma a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL”

(Reglamento de Graduación de la ESPOL)

Freddy Samuel Navarro Tobar

RESUMEN

En la actualidad el banano se encuentra afectado por muchas enfermedades, en esta tesis estudiaremos la enfermedad foliar que tiene mayor incidencia a nivel mundial esta es la Sigatoka Negra. Analizando a través de procedimientos estadísticos multivariados los datos recopilados en la Hacienda Carmita en Balao provincia del Guayas a través del tiempo año 2005 y 2006, verificaremos la relación que existe entre la enfermedad y los parámetros relacionados con el clima, estos datos son de fuente exclusiva del CIBE (Centro de Investigación Biotecnológicas del Ecuador).

Entre las variables a estudiar tenemos tres grupos: Preaviso son las variables II, III, IV que nos indican el estado de la enfermedad en la Hoja II, Hoja III y Hoja IV; Severidad son las variables Floración y Cosecha que nos indican el estado en el que se encuentra la enfermedad en su etapa de Floración y Cosecha; y el grupo Clima son las variables Rain, Temp Out, Temp Hi, Wind Chill, Heat Index, Solar Rad y otras.

En el primer capítulo presentamos la reseña histórica del banano, la superficie bananera en el Ecuador, los principales mercados, sus caracteres botánicos como está clasificado y cuales son sus prácticas agronómicas.

En el segundo capítulo trataremos sobre la enfermedad foliar la Sigatoka donde es su origen, sus antecedentes, cuales son las medidas de exclusión y la evaluación de la enfermedad.

En el tercer capítulo conoceremos la estructura, las fórmulas y los métodos estadísticos univariados y multivariados a utilizar.

En el cuarto capítulo se realiza el análisis univariado y multivariado de los datos de las variables para ver cual es la relación que existe entre la enfermedad y los parámetros de clima.

INDICE GENERAL

	Pág.
RESUMEN	II
INDICE GENERAL	IV
ABREVIATURAS	X
SIMBOLOGIA	XI
INDICE DE FIGURAS	XII
INDICE DE TABLAS	XVI
CAPÍTULO I	
1. El Banano	1
1.1. Introducción	2
1.2. El Banano en el Ecuador	5
1.3. Superficie Bananera en el Ecuador	8
1.4. Principales Mercados	9
1.5. Caracteres Botánicos	9
1.5.1. Rizoma o Bulbo	9
1.5.2. Sistema Radicular	10
1.5.3. Sistema Foliar	10
1.5.4. Inflorescencia	11
1.5.5. El Fruto	12
1.6. Variedades	13
1.6.1. Grosh Michel	13
1.6.2. Cavendish	14
1.6.2.1. Lacatan o Filipino	14
1.6.2.2. Poyo	15
1.6.2.3. Giant Cavendish	15
1.6.2.4. Cavendish Enano	15
1.7. Requerimientos del Cultivo	16
1.7.1. Requerimientos Ecológicos	16

1.7.1.1. El Suelo	16
1.7.1.2. El Clima	16
1.7.2. Selección del Terreno	17
1.7.3. Levantamiento Topográfico.....	17
1.7.4. Obras Previas a la Siembra	18
1.7.4.1. Preparación del Terreno	18
1.7.4.2. Canales de Riego	18
1.7.4.3. Canales de Drenaje	18
1.7.5. Siembra	19
1.7.5.1. Densidades de Siembre	19
1.7.5.2. Selección de Semilla	20
1.7.5.3. Preparación de Semilla	20
1.7.5.4. Alineación y Estaquillado	21
1.7.5.5. Forma de Siembra	21
1.7.6. Prácticas Agronómicas	21
1.7.6.1. Riego	21
1.7.6.2. Control de Malezas	22
1.7.6.3. Fertilización	22
1.7.6.4. Deshije	23
1.7.6.5. Deshoje	25
1.7.6.6. Apuntalado	25
1.7.6.7. Enfunde	25
1.7.6.8. Desmane	26
1.7.7. Enfermedades del Banano	26
CAPÍTULO II	
2. La Sigatoka	30
2.1. Introducción	31
2.2. Antecedentes	34
2.2.1. Estado 1	37
2.2.2. Estado 2	37

2.2.3. Estado 3	37
2.2.4. Estado 4	38
2.2.5. Estado 5	38
2.2.6. Estado 6	39
2.3. Medidas de Exclusión	40
2.4. Evaluación de la Enfermedad	43
2.4.1. Preaviso	43
2.4.2. Severidad	44
CAPÍTULO III	
3. Métodos Estadísticos	45
3.1. Introducción	46
3.2. Definiciones Básicas	46
3.2.1. Población	46
3.2.2. Parámetro	47
3.2.3. Muestra	47
3.2.4. Estadístico	47
3.2.5. Variable	47
3.2.6. Muestreo	48
3.2.7. Error de Muestreo	49
3.2.8. Sesgo Muestral	50
3.2.9. Medidas de Tendencia Central	50
3.2.9.1. Media	50
3.2.9.1.1. Media Poblacional	50
3.2.9.1.2. Media Muestral	51
3.2.9.2. Mediana	51
3.2.10. Medidas de Dispersión	52
3.2.10.1. Rango	52
3.2.10.2. Varianza y Desviación Estándar Poblacional	52
3.2.10.3. Varianza y Desviación Estándar Muestral	53
3.2.10.4. Otras Medidas de Dispersión	54

3.2.10.4.1. Deciles, Cuartiles y Percentiles	54
3.2.10.4.2. Rango o Recorrido Intercuatico.....	56
3.3. Uso Frecuente de la Desviación Estándar	57
3.3.1. Distribución Normal	57
3.3.2. Regla Empírica	57
3.3.3. Sesgo	58
3.4. Medidas de Distribución	59
3.4.1. Asimetría	59
3.4.2. Curtosis	61
3.5. Coeficientes de Variación	63
3.6. Figuras de Cajas (Box - Plot)	64
3.7. Biplots	64
3.8. Imputación de Datos	67
3.8.1. Imputación por Medias	69
3.8.2. Imputación por Regresión Polinómicas	69
3.9. Bondad de Ajuste (Kolmogorov)	70
3.10. Coeficiente de Correlación de Pearson	72
3.11. Distribución Ji Cuadrada	73
3.12. Distribución F de Fisher	73
3.13. Análisis de Conglomerados	74
3.14. Vector Medio Muestral Multivariada	76
3.15. Análisis de Componentes Principales	78
3.16. Análisis de Correspondencias	84
CAPÍTULO IV	
4. Análisis Estadístico	88
4.1. Introducción	89
4.2. Análisis de Variables	89
4.3. Análisis Univariado	90
4.3.1. Variable Preaviso	91
4.3.1.1. Hoja III	91

4.3.1.2. Hoja IV	93
4.3.1.3. Estado Evolutivo	94
4.3.2. Severidad	96
4.3.2.1. Floración	96
4.3.2.2. Cosecha	98
4.3.3. Clima	100
4.3.3.1. Rain Acumulada.....	100
4.3.3.2. Rain Promedio	102
4.3.3.3. Tem Out Acumulada	104
4.3.3.4. Temp Out Promedio	106
4.3.3.5. Temp Hi Acumulada	108
4.3.3.6. Temp Hi Promedio	110
4.3.3.7. Temp Low Acumulada	112
4.3.3.8. Temp Low Promedio	114
4.3.3.9. Wind Chill Acumulada	116
4.3.3.10. Wind Chill Promedio	118
4.3.3.11. Heat Index Acumulada.....	120
4.3.3.12. Heat Index Promedio	122
4.3.3.13. THW Index Acumulada	124
4.3.3.14. THW Index Promedio	126
4.3.3.15. THWS Index Acumulada	128
4.3.3.16. THWS Index Promedio	130
4.3.3.17. Cool D-D Acumulada.....	132
4.3.3.18. Cool D-D Promedio	134
4.3.3.19. Solar Rad Acumulada	136
4.3.3.20. Solar Rad Promedio	138
4.3.3.21. Solar Energy Acumulada	140
4.3.3.22. Solar Energy Promedio	142
4.3.3.23. Hi Solar Rad Acumulada.....	144
4.3.3.24. Hi Solar Rad Promedio	146

4.3.3.25. UV Index Acumulada	148
4.3.3.26. UV Index Promedio	150
4.3.3.27. UV Dose Acumulada	152
4.3.3.28. UV Dose Promedio	154
4.3.3.29. Hi UV Acumulada	156
4.3.3.30. Hi UV Promedio	158
4.3.3.31. ET Acumulada	160
4.3.3.32. ET Promedio	162
4.3.3.33. Out Hum Acumulada	164
4.3.3.34. Out Hum Promedio	166
4.3.3.35. Dew Pt. Acumulada	168
4.3.3.36. Dew Pt. Promedio	170
4.3.3.37. Wind Speed Acumulada	172
4.3.3.38. Wind Speed Promedio	174
4.4. Bondad de Ajuste (Kolmogorov)	176
4.5. Análisis de Correlación de Pearson	178
4.6. Análisis de Conglomerados	180
4.7. Componentes Principales	182
4.8. Análisis de Correspondencia	186
CAPÍTULO V	
5. Conclusiones y Recomendaciones	190
5.1. Conclusiones	191
5.2. Recomendaciones	194

ABREVIATURAS

(°)	Grados
C	Centígrados
mts.	Metros
ctms.	Centímetros
mm.	Milímetros
w/m ²	Watios por metros cuadrados
Has.	Hectáreas
%	Porcentaje
gl.	Grados de Libertad
M.	Mycosphaerella
III	Hoja III
IV	Hoja IV
EE	Estado Evolutivo
Fig.	Figura
IC	Intervalo de Confianza
ACP	Análisis de Componentes Principales
AC	Análisis de Correspondencia
ACS	Análisis de Correspondencia Simple
Q ₁	Primer Cuartil
Q ₂	Segundo Cuartil
Q ₃	Tercer Cuartil

SIMBOLOGIA

$^{\circ}\text{C}$	Grados centígrados
X	Variable independiente
Y	Variable dependiente
N	Número de observaciones de una población
n	Número de observaciones de una muestra
μ	Media Poblacional
\bar{x}	Media Muestral
σ^2	Varianza Poblacional
s^2	Varianza Muestral
λ	Valores Propios
H_0	Hipótesis Nula
H_1	Hipótesis Alternativa
β	Coefficiente de una variable
β_i	Coefficiente de la variable i-ésima
ε	Error de la observación
X'	Transpuesta de una matriz
$f(\chi^2)$	Función de densidad Ji Cuadrado

ÍNDICE DE FIGURAS

		Pág.
Figura 2.1	Porcentaje de clasificación de la Severidad	44
Figura 3.1	Recorrido Intercuatico	56
Figura 3.2	Curvas Sesgadas	58
Figura 3.3	Tipos de Curvas Asimétricas	60
Figura 3.4	Tipos de Curtosis	61
Figura 3.5	Curva Normal	63
Figura 4.1	Diagramas de Cajas del Estado Evolutivo de la Enfermedad en la Hoja III.	91
Figura 4.2	Histograma del Estado Evolutivo de la Enfermedad de Hoja II.	92
Figura 4.3	Diagramas de Cajas del Estado Evolutivo de la Enfermedad en la Hoja IV.	93
Figura 4.4	Histograma del Estado Evolutivo de la Enfermedad de Hoja IV.	94
Figura 4.5	Diagramas de Cajas del Estado Evolutivo de la Enfermedad en la Hoja IV.	95
Figura 4.6	Histograma del Estado Evolutivo	96
Figura 4.7	Diagrama de Cajas de la Enfermedad en la Floración	97
Figura 4.8	Histograma del Estado de la Enfermedad en la Floración.	98
Figura 4.9	Diagrama de Cajas del Estado de la Enfermedad de Cosecha.	99
Figura 4.10	Histograma del Estado de la Enfermedad en la Cosecha.	100
Figura 4.11	Diagrama de Cajas de Rain Acumulada	101
Figura 4.12	Histograma de Rain Acumulada	102
Figura 4.13	Diagrama de Cajas de Rain Promedio	103
Figura 4.14	Histograma de Rain Promedio	104

Figura 4.15	Diagrama de Cajas de Temp Out Acumulada	105
Figura 4.16	Histograma de Temp Out Acumulada	106
Figura 4.17	Diagrama de Cajas de Temp Out Promedio	107
Figura 4.18	Histograma de Temp Out Promedio	108
Figura 4.19	Diagrama de Cajas de Temp Hi Acumulada	109
Figura 4.20	Histograma de Temp Hi Acumulada	110
Figura 4.21	Diagrama de Cajas de Temp Hi Promedio	111
Figura 4.22	Histograma de Temp Hi Promedio	112
Figura 4.23	Diagrama de Cajas de Temp Low Acumulada	113
Figura 4.24	Histograma de Temp Low Acumulada	114
Figura 4.25	Diagrama de Cajas de Temp Low Promedio	115
Figura 4.26	Histograma de Temp Low Promedio	116
Figura 4.27	Diagrama de Cajas de Wind Chill Acumulada	117
Figura 4.28	Histograma de Wind Chill Acumulada	118
Figura 4.29	Diagrama de Cajas de Wind Chill Promedio	119
Figura 4.30	Histograma de Wind Chill Promedio	120
Figura 4.31	Diagrama de Cajas de Heat Index Acumulada	121
Figura 4.32	Histograma de Heat Index Acumulada	122
Figura 4.33	Diagrama de Cajas de Heat Index Promedio	123
Figura 4.34	Histograma de Heat Index Promedio	124
Figura 4.35	Diagrama de Cajas de THW Index Acumulada	125
Figura 4.36	Histograma de THW Index Acumulada	126
Figura 4.37	Diagrama de Cajas de THW Index Promedio	127
Figura 4.38	Histograma de THW Index Promedio	128
Figura 4.39	Diagrama de Cajas THSW Index Acumulada	129
Figura 4.40	Histograma de THSW Index Acumulada	130
Figura 4.41	Diagrama de Cajas de THSW Index Promedio	131
Figura 4.42	Histograma de THSW Index Promedio	132
Figura 4.43	Diagrama de Cajas de Cool D-D Acumulada	133
Figura 4.44	Histograma de Cool D-D Acumulada	134

Figura 4.45	Diagrama de Cajas de Cool D-D Promedio	135
Figura 4.46	Histograma de Cool D-D Promedio	136
Figura 4.47	Diagrama de Cajas de Solar Rad Acumulada	137
Figura 4.48	Histograma de Solar Rad Acumulada	138
Figura 4.49	Diagrama de Cajas Solar Rad Promedio	139
Figura 4.50	Histograma de Solar Rad Promedio	140
Figura 4.51	Diagrama de Cajas Solar Energy Acumulada	141
Figura 4.52	Histograma de Solar Energy Acumulada	142
Figura 4.53	Diagrama de Cajas de Solar Energy Promedio	143
Figura 4.54	Histograma de Solar Energy Promedio	144
Figura 4.55	Diagrama de Cajas de Hi Solar Rad Acumulada	145
Figura 4.56	Histograma de Hi Solar Rad Acumulada	146
Figura 4.57	Diagrama de Cajas de Hi Solar Rad Promedio	147
Figura 4.58	Histograma de Hi Solar Rad Promedio	148
Figura 4.59	Diagrama de Cajas UV Index Acumulada	149
Figura 4.60	Histograma de UV Index Acumulada	150
Figura 4.61	Diagrama de Cajas de UV Index Promedio	151
Figura 4.62	Histograma de UV Index Promedio	152
Figura 4.63	Diagrama de Cajas de UV Dose Acumulada	153
Figura 4.64	Histograma de UV Dose Acumulada	154
Figura 4.65	Diagrama de Cajas de UV Dose Promedio	155
Figura 4.66	Histograma de UV Dose Promedio	156
Figura 4.67	Diagrama de Cajas de Hi UV Acumulada	157
Figura 4.68	Histograma de Hi UV Acumulada	158
Figura 4.69	Diagrama de Cajas de Hi UV Promedio	159
Figura 4.70	Histograma de Hi UV Promedio	160
Figura 4.71	Diagrama de Cajas de ET Acumulada	161
Figura 4.72	Histograma de ET Acumulada	162
Figura 4.73	Diagrama de Cajas de ET Promedio	163
Figura 4.74	Histograma de ET Promedio	164

Figura 4.75	Diagrama de Cajas de Out Hum Acumulada	165
Figura 4.76	Histograma de Out Hum Acumulada	166
Figura 4.77	Diagrama de Cajas de Out Hum Promedio	167
Figura 4.78	Histograma de Out Hum Promedio	168
Figura 4.79	Diagrama de Cajas de Dew Pt. Acumulada	169
Figura 4.80	Histograma de Dew Pt. Acumulada	170
Figura 4.81	Diagrama de Cajas de Dew Pt. Promedio	171
Figura 4.82	Histograma de Dew Pt. Promedio	172
Figura 4.83	Diagrama de Cajas de Wind Speed Acumulada	173
Figura 4.84	Histograma de Wind Speed Acumulada	174
Figura 4.85	Diagramas de Cajas de Wind Speed Promedio	175
Figura 4.86	Histograma de Wind Speed Promedio	176
Figura 4.87	Correlación entre Variable III y IV	179
Figura 4.88	Análisis de Conglomerados Variables de Infección.....	181
Figura 4.89	BIPLOT de Componentes Principales	183
Figura 4.90	Análisis de Correspondencia	189

ÍNDICE DE TABLAS

		Pág.
Tabla I	Distancia entre Grupos	181
Tabla II	Auto Vectores	184
Tabla III	Autovalores	185
Tabla IV	Cuartiles para Categorización	186
Tabla V	Ji Cuadrado	188

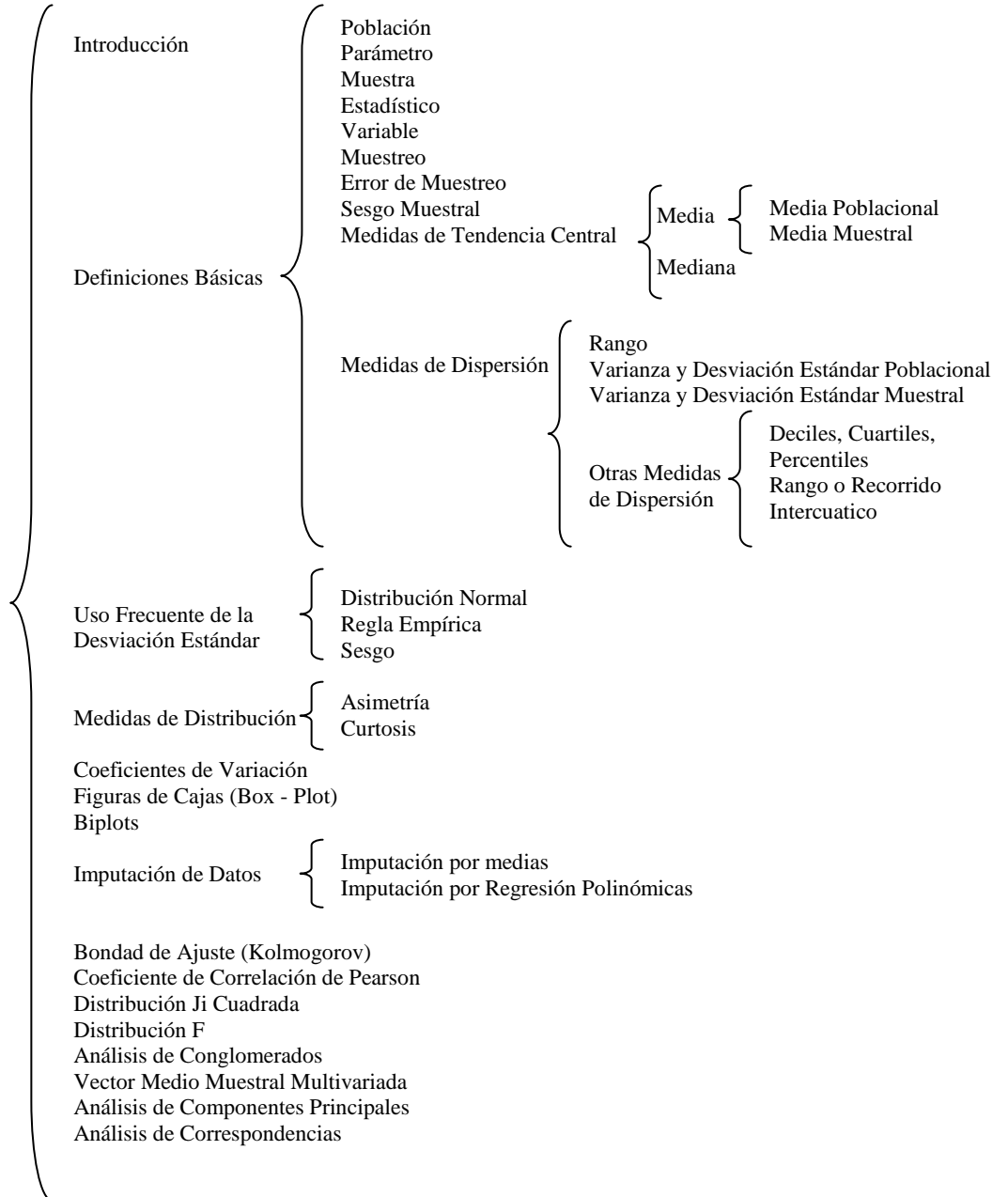
**CAPÍTULO I
EL BANANO**

Introducción		
El Banano en el Ecuador		
Superficie Bananera en el Ecuador		
Principales Mercados		
Caracteres Botánicos	{	Rizoma o Bulbo Sistema Radicular Sistema Foliar Inflorescencia El Fruto
Variedades	{	Grosh Michel Cavendish
	{	Lacatan o Filipino Poyo Giant Cavendish Cavendish Enano
Requerimientos del Cultivo	{	Requerimientos Ecológicos { El suelo El clima Selección del Terreno { Preparación del Terreno Levantamiento Topofigura { Canales de Riego Obras previas a la Siembra { Canales de Drenaje
Enfermedades del Banano	{	Siembra { Densidades de siembra Selección de semilla Preparación de semilla Alineación y estaquillado Forma de siembra
	{	Prácticas Agronómicas { Riego Control de Malezas Fertilización Deshije Deshoje Apuntalado Enfunde Desmane

**CAPÍTULO II
LA
SIGATOKA**

Introducción	{	Estado 1 Estado 2 Estado 3 Estado 4 Estado 5 Estado 6
Antecedentes	{	
Medidas de Exclusión	{	
Evaluación de la Enfermedad	{	Preaviso Severidad

**CAPÍTULO III
MÉTODOS
ESTADÍSTICOS**



**CAPÍTULO IV
ANÁLISIS
ESTADÍSTICO**

Introducción
Análisis de Variables

Preaviso

Hoja III
Hoja IV
Estado Evolutivo

Severidad

Floración
Cosecha

Análisis Univariado

Clima

Rain Acumulada
Rain Promedio
Temp Out Acumulada
Temp Out Promedio
Temp Hi Acumulada
Temp Hi Promedio
Temp Low Acumulada
Temp Low Promedio
Wind Chill Acumulada
Wind Chill Promedio
Heat Index Acumulada
Heat Index Promedio
THW Index Acumulada
THW Index Promedio
THWS Index Acumulada
THWS Index Promedio
Cool D-D Acumulada
Cool D-D Promedio
Solar Rad Acumulada
Solar Rad Promedio
Solar Energy Acumulada
Solar Energy Promedio
Hi Solar Rad Acumulada
Hi Solar Rad Promedio
UV Index Acumulada
UV Index Promedio
UV Dose Acumulada
UV Dose Promedio
Hi UV Acumulada
Hi UV Promedio
ET Acumulada
ET Promedio
Out Hum Acumulada
Out Hum Promedio
Dew Pt. Acumulada
Dew Pt. Promedio
Wind Speed Acumulada
Wind Speed Promedio

Bondad de Ajuste (Kolmogorov)
Análisis de Correlación de Pearson
Análisis de Conglomerados
Componentes Principales
Análisis de Correspondencia

**CAPÍTULO V
CONCLUSIONES Y
RECOMENDACIONES**

Conclusiones
Recomendaciones

CAPÍTULO I

EL BANANO

INTRODUCCIÓN

El Banano es una planta herbácea gigante, pertenece al género *Musa*, familias de las Musáceas; posee algunas especies como ***Musa sapientum***, ***M. paradisíaca***, ***M. textilis***, ***M. ornamental***, etc., de las cuales, las dos primeras son las más cultivadas en nuestro país, sin desconocer que la *M. textilis* también es un producto de exportación.

El Banano comestible se originó a través de una serie de mutaciones y cambios genéticos, a partir de especies silvestres no comestibles, de fruto pequeño con numerosas semillas. Para llegar a las mutaciones se producen cambios en los cromosomas que tienen las características hereditarias que dieron origen al banano comestible comercial.

Los orígenes del banano se pierden en la noche de los tiempos. Nos remontamos a miles de años atrás y vemos que hallazgo arqueológico de algunos fósiles revela la existencia del banano miles de años antes de Cristo.

Posiblemente el banano es la planta cultivada más antigua en el mundo. Hace más de tres mil años, ya se lo mencionaba en escritos chinos como uno de los primeros alimentos del hombre primitivo. Los antiguos lo llamaban “fruta de los hombres sabios” dada sus cualidades nutricionales superiores.

El profesor Wardlaw sostiene que el banano es originario de las zonas tropicales comprendidas entre Indo-Malaya y las islas más al Sur. Paúl G. Minneman cree, en cambio, que el banano es originario de América como el caucho y fue llevado al otro lado del hemisferio y desarrollado su cultivo en el Sur de Asia.

Un “manjar de los Dioses” se acostumbra decir en el argot culinario, y exclusivamente asiático hasta el siglo VII, cuando es llevado a África Occidental por los mercaderes árabes, dónde habría de recibir su nombre definitivo, formado por la contracción de palabras del continente negro, como son Bana, gbaná, abana, funana y bunane.

El banano no se radica exclusivamente en África Occidental, sino que iría a las Islas Canarias, llevada por exploradores y sacerdotes portugueses y españoles de la época, estableciéndose por estos territorios. Con el descubrimiento y colonización de América, en 1516, Fray Tomás de Berlanga, descubridor de las islas Galápagos trajo estas plantas a R. Dominicana y de allí se dispersan al continente, incluyendo el Ecuador de hoy.

A los pueblos de América Latina llegó el banano a inicios del siglo XVI, traído por Fray Tomás de Berlanga, que lo sembró concretamente en la isla Española y de allí se esparció a los demás pueblos de la región, sirviendo como una fruta exótica pero muy nutritiva, que se destinaba al consumo de sus habitantes.

La explotación comercial del banano arranca en el área del Caribe, cuando un químico francés, Jean Francois Pouyat, que vivía en la Isla de Jamaica encontró la especie gross michel en Martinica, sin saber cómo llegó, siendo posible que haya sido originaria, tanto de América como del sur de Asia.

Cuando Pouyat regresó a Jamaica, llevó consigo unas cuantas cepas y colines que sembró en la isla, donde se formaron las primeras bananeras. Por esto Jamaica le entregó un reconocimiento, consistente en una moneda de oro, al cosechar los primeros racimos de banano.

Entre los años 1800 y 1849, cuando se iniciaron las primeras y ocasionales remesas de la fruta, desde los países del Caribe hacia los Estados Unidos de Norte América, y en pequeños volúmenes, a través de marinos que realizaban viajes de cabotaje cubriendo la ruta hacia El Caribe, Centroamérica y América del Sur.

Pero sería en julio de 1876, con oportunidad de la Exposición del Centenario de Filadelfia, cuando llegaron los primeros embarques de la fruta, procedentes de Centroamérica, congregándose la muchedumbre en los muelles de Nueva York y Boston, maravillándose con el sabor de la fruta, que adquirió de esta manera su carta de naturalización en el mercado norteamericano.

M. paradisíaca y *M. sapientum* fueron las primeras especies introducidas a América. Las variedades Gros Michel y Cavendish se introdujeron a comienzos del siglo XIX, no hay fecha de indicios de la introducción al Ecuador de la planta de banano.

1.2. EL BANANO EN EL ECUADOR

El Instituto de Investigaciones Económicas y Políticas de la Universidad de Guayaquil da a conocer que los primeros registros de la actividad bananera en el país, se remontan a las exportaciones, desde 1910, con facturaciones en racimos hacia los mercados de Perú, Chile y Estados Unidos.

Con la creación de la Asociación Nacional de Bananeros del Ecuador (ANBE) en junio de 1955, se inicia una recopilación de registros de superficie, exportación, destino, incautación de divisas que son recuperados en 1992,

en las estadísticas de exportación de la fruta de ese año, por el Programa Nacional del Banano y Frutas Tropicales.

De la revisión de las estadísticas de ambas instituciones, obtenemos una lectura que nos demuestra, un ritmo creciente en las exportaciones de bananos del Ecuador, desde inicios del siglo anterior, pero debiendo experimentar, tres períodos en los que las ventas de la fruta hacia el mercado internacional, declinan de manera considerable, por las dos guerras mundiales de 1914 a 1918 y de 1942 a 1945; y por la depresión de los Estados Unidos en 1932.

Y es que terminado el primer conflicto bélico mundial, se experimenta un notable repunte en las exportaciones de la fruta, que luego se afecta con la depresión económica que soporta Estados Unidos en 1932 y que se refleja en la caída de las exportaciones de la fruta, a 324.387 racimos en ese año, que repunta para el año 1936 con la venta de 2 millones 220 mil 513 racimos.

Entre 1942 y 1945, por la segunda guerra mundial, se derrumba la exportación de bananos, en forma drástica nuevamente, para recuperarse en el año 1946, con ventas por un millón 301 mil 409 racimos. Superado el conflicto, se dará un repunte en las ventas al mercado mundial el año 1951, con 16 millones 755 mil 66 racimos, que generaron 320 millones 773 mil 951

sucres, según el Boletín de Investigaciones Económicas y Políticas de la Universidad de Guayaquil.

El comercio internacional de la fruta, ubica a los países centroamericanos como los primeros vendedores hacia el mercado de Estados Unidos, por marinos mercantes del gran país del norte en medio de una gran aceptación, que generaliza la siembra comercial, por inversionistas que se dedican al tendido de líneas férreas y venta de locomotoras, a cambio de la concesión de tierras para la siembra de bananos, que alumbrarán las compañías transnacionales.

El Ecuador desde 1952 es el primer exportador de banano al mundo, con la única excepción de los años 1982 y 1983, cuando por efectos de un fenómeno natural de El Niño, soportó el arrasamiento de gran parte de sus plantaciones, cediéndole este liderazgo a Costa Rica, segundo exportador a nivel mundial, pero en 1984 volvió a ocupar el liderazgo que lo mantiene, gracias a la forma de combinar los recursos del trabajo, técnica, inversión económica, administración y tierra.

1.3. SUPERFICIE BANANERA EN EL ECUADOR.

Según datos oficiales del Ministerio de Agricultura y Ganadería, tomado a diciembre 31 del 2003, las áreas de El Oro, Guayas y Los Ríos (las principales del país) y de otras provincias inscritas en dicha dependencia, ascienden a 153.403 hectáreas, desglosadas de la siguiente manera:

Los Ríos (Has.) 49.246,13

El Oro (Has.) 45.839,04

Guayas (Has.) 45.503,36

Otras (Has.) 12.814,47

TOTAL (Has.) 153.403,83

Hoy, consideramos que a más del área inscrita, la superficie cultivada incluida la no registrada asciende estimativamente a 180 mil hectáreas, según la siguiente composición:

Los Ríos 55.000 has.

Guayas 50.000 has.

El Oro 48.000 has.

Cañar 8.000 has

Cotopaxi 4.000 has.

Otras provincias: (Bolívar, Esmeralda, Pichincha, Manabí) 15.000 has.

TOTAL 180.000 Has.

1.4. PRINCIPALES MERCADOS

Los principales mercados para el banano Ecuatoriano son: Estados Unidos, Unión Europea, Europa del Este, Cono Sur, Oriente, Medio Oriente, África del Norte, Oceanía.

1.5. CARACTERES BOTÁNICOS

1.5.1. RIZOMA O BULBO

Llamado comúnmente cepa, produce una yema vegetativa que sale de la planta madre y sufre un cambio anatómico y morfológico de los tejidos y al crecer diametralmente forma el rizoma que alcanza una considerable altura.

Al dar origen a la planta, en la zona interna se originan las raíces y yemas vegetativas que serán los nuevos retoños o hijos. Cada planta nace en forma de brote y crece en la base de la planta madre o tallo principal de la cual depende para su nutrición hasta cuando produce hojas anchas y se autoabastece.

1.5.2. SISTEMA RADICULAR

Posee raíces superficiales se distribuye en una capa de 30 a 40 cts. y se encuentra mayor concentración de raíces en la capa de 15 a 20 cts.

Las raíces son de color blanco, tiernas cuando emergen y se vuelven amarillentas y duras, su diámetro oscila entre 5 y 8 mm., la longitud varía y puede llegar de 2,5 a 3 mts. de crecimiento lateral y hasta 1,5 mts. de profundidad.

El poder de penetración de las raíces del banano es débil, su distribución radicular está relacionada con la textura y estructura del suelo.

1.5.3. SISTEMA FOLIAR

Las hojas del banano se originan del punto central de crecimiento o meristema terminal, situado en la parte superior del Bulbo. Luego se nota precozmente la formación del pecíolo y la nervadura central terminada en filamento, lo que será la vaina posteriormente. La parte de la nervadura se alarga y el borde izquierdo comienza a cubrir el borde derecho, los cuales crecen en altura y forman los semilimbos. La hoja se forma en el interior del pseudotallo.

La hoja emerge enrollada en forma de cigarro. Una vez que ha salido la tercera parte de la longitud, la presencia de la coloración verde o pigmentación clorofílica se hace inmediatamente.

La producción de las hojas cesa cuando emerge la inflorescencia, o sea cuando la planta “pare”.

Una hoja adulta completa está formada por las siguientes partes: vaina, pecilio, nervadura, limbo.

Las dimensiones de los limbos varían de 70 a 100 centímetros de ancho por 3 ó 4 mts. de largo, su espesor de 0.35 a 1mm., tiene una cantidad de 11 a 12 hojas, al momento de la emisión floral.

1.5.4. INPLORESCENCIA

La yema floral es corta y cónica, este cambio en el punto de crecimiento marca el comienzo del crecimiento del tallo verdadero que ha permanecido a ras del suelo y se convertirá en un tallo aéreo y crecerá por el centro del pseudotallo.

Las células de la yema floral continuarán creciendo longitudinalmente y hacia arriba por la parte central del pseudotallo para emerger por la parte superior de la planta.

Durante el crecimiento dentro del pseudotallo los brotes florales se diferencian y principian su desarrollo al emerger la bellota o inflorescencia, ya están diferenciados los brotes florales con el número de dedos y manos.

Las flores femeninas y las masculinas quedan expuestas, las flores femeninas agrupadas en grupos de dos filas apretadas y sobrepuestas, lo que se conoce con el nombre de mano y su distribución está en forma helicoidal a lo largo del eje floral; al conjunto de flores femeninas agrupadas en manos se lo conocen con el nombre de "racimo".

1.5.5. EL FRUTO

Se desarrolla de los ovarios de las flores pistiladas por el aumento del volumen de las tres celdas del ovario, opuestas al eje central. Los ovarios abortan y salen al mismo tiempo los tejidos del pericarpio o cáscara y engrosan, la actividad de los canales de látex disminuye, cesando por completo cuando el fruto está maduro.

La parte comestible que resulta del engrosamiento de las paredes del ovario, es una masa de parénquima cargada de azúcar y almidón, en la madurez no hay células activas de taninos, ni tejidos fibrosos.

Los tres lóbulos que forman el ovario se pueden separar longitudinalmente por sus planos de unión. En el lóbulo inmediato a la cáscara se encuentra un surco fino longitudinal que corresponde a cada una de las haces vasculares principales.

En un corte transversal aparecen muchos haces vasculares como puntos de color más claro sobre el fondo blanco del parénquima y del endocarpio que está presentado por paredes de células delgadas radiales, que en la madurez permiten separar la cáscara de la parte central de la fruta.

1.6. VARIEDADES

En el Ecuador se cultiva las variedades como: gross michel y cavendish.

1.6.1. GROS MICHEL.- Es una variedad grande y robusta, su pseudotallo tiene una longitud de 6 u 8 mts. de coloración verde claro con tono a rosa por algunas partes.

Su peciolo en la base con manchas de color marrón oscuro. Los limbos son verdes definidos siendo de 4 mts. de largo por 1 m de ancho. Los racimos penden verticalmente, son alargados de forma cilíndrica con 10 a 14 manos promedio. Los frutos de la fila interna se muestran erectos pues su curva se encuentra en el pedúnculo y en la parte basal del fruto.

El ápice tiene forma de cuello de botella, el pedúnculo es más corto y robusto. La maduración es regular y homogénea de sabor delicioso y textura delicada por lo que se le conoce con el nombre de “guineo de seda” o “plátano de seda”.

1.6.2. CAVENDISH

Tiene 7 tipos principales: lacatan o filipino, poyo, giant cavendish, cavendish enano, valery, robusta, grand nain, de los cuales enunciaremos cuatro.

1.6.2.1. LACATAN o FILIPINO

Alcanza alturas de 4 a 6 mts. con racimos largos de forma cilíndrica, los frutos son curvados en la parte apical. Los pedúnculos son largos y frágiles, el fruto muy sensible a los parásitos después de cosechados, la maduración es delicada y menos atractiva.

1.6.2.2. POYO

Llamado Robusto tiene una altura entre 2,8 mts. y 4 mts. lo que le hace más resistente al viento, sus hojas son más cortas, son de rápido desarrollo y se puede sembrar a mayor densidad, por lo que aumenta su productividad. Su racimo es compacto por tener sus entrenudos cortos y sus frutos son rectos y largos.

1.6.2.3. EL GIANT CAVENDISH

Tiene una altura entre 2,8 mts. y 3 mts. sus limbos son cortos y grandes con falsos entrenudos cortos. Su racimo es de aspecto troncocónico. Las dos primeras manos se mantienen rectas pero están curvadas en las demás manos. Es una variedad muy productiva y la conformación de sus racimos es muy buena.

1.6.2.4. CAVENDISH ENANO

Posee un pseudotallo con altura de 2 mts. sus limbos foliares son anchos y cortos. Los racimos presentan una forma troncocónica definida porque sus manos están más cerca unas con otras y abiertas en relación al eje, los frutos son curvos.

1.7. REQUERIMIENTOS DEL CULTIVO

1.7.1. REQUERIMIENTOS ECOLOGICOS

La planta de banano crece en las más variadas condiciones del suelo y clima; es necesario tomar en cuentas las condiciones más favorables y son:

1.7.1.1. SUELO.- Los suelos aptos para el desarrollo del cultivo de banano son aquellos que presentan una textura: franco arenosa, franco arcillosa, franco arcillo limoso y franco limoso; además deben poseer un buen drenaje interno y alta fertilidad, su profundidad debe ser de 1,2 a 1,5 mts.

Por otro lado deben poseer buenas propiedades de retención de agua, los suelos arcillosos con un 40% no son recomendables para el cultivo.

1.7.1.2. CLIMA.- El clima ideal es el tropical húmedo. La temperatura adecuada va desde los 18,5°C a 35,5°C. A temperaturas inferiores de 15,5°C se retarda el crecimiento.

Con temperaturas de 40°C no se han observado efectos negativos siempre y cuando la provisión de agua sea normal.

La pluviosidad necesaria varía de 120 a 150mm de lluvia mensual o precipitaciones de 44mm semanales. En nuestro Litoral Ecuatoriano es necesario realizar el riego porque tiene definido sus estaciones lluviosa y seca. El banano requiere de buena luminosidad y ausencia de vientos fuertes.

1.7.2. SELECCIÓN DEL TERRENO.- Un terreno para cultivo de banano debe tomarse en cuenta el clima, el suelo, las vías de comunicación que posee, las condiciones de las vías, la facilidad de obtener y transportar agua de riego, qué cultivos se sembraron anteriormente, qué pesticidas se utilizaron, la topografía y otros factores que podrían eliminar la producción de la fruta.

1.7.3. LEVANTAMIENTO TOPOGRÁFICO.- Es de vital importancia tener un levantamiento topográfico del terreno para realizar el cultivo del banano, para la ubicación de canales de drenaje, canales de riego o tendido de tuberías, los sitios donde se construirán las empacadoras, el diseño de ubicación de funiculares y cables vías, ubicación de guardarrayas otro trabajo que requiere el cultivo.

1.7.4. OBRAS PREVIAS A LA SIEMBRA

1.7.4.1. PREPARACIÓN DEL TERRENO.- Se eliminarán todos los obstáculos del terreno, se procede a arar y rastrar hasta conseguir buena uniformidad del suelo, así como una buena aireación.

1.7.4.2. CANALES DE RIEGO.- Se realiza la distribución de los canales de riego, sean éstos primarios o secundarios, así como la ubicación de compuertas y tomas de agua.

Si el sistema de riego a utilizarse es por aspersión, se hará la distribución de las moto bombas, tuberías, torres, se debe tomar en cuenta las necesidades del cultivo.

1.7.4.3. CANALES DE DRENAJE.- Es muy importante los canales de drenaje para bajar el nivel freático o tabla de agua, o para eliminar el agua superficial de los pozos o charcas.

La profundidad de los canales de drenajes está determinada por las propiedades físicas del suelo, la intensidad y frecuencias de las lluvias; por lo general deben tener una profundidad de 1,20 y 1,50 mts.

1.7.5. SIEMBRA

1.7.5.1. DENSIDADES DE SIEMBRA.- Gran parte del rendimiento del cultivo depende de la “densidad poblacional” ideal para una región, se deben tomar en cuenta los siguientes parámetros:

- a. **Variedad:** Mientras más pequeña sea, la densidad será mayor.
- b. **Lluvia:** Es importante la precipitación para determinar la densidad de siembra; por ejemplo, cuando hay mayor precipitación la densidad será mayor, cuando hay menos precipitación la densidad será menor.
- c. **Propiedades físicas y químicas del suelo:** En suelos livianos menor densidad, en suelos pesados mayor densidad, en suelos fértiles menor densidad, en suelos menos fértiles mayor densidad.
- d. **Sistema de deshije:** Mediante ésta práctica se determina la población efectiva dejando el número de plantas adecuado por unidad de superficie.

La densidad de siembra varía de acuerdo a la variedad, tipo de suelo, tipo de drenaje, sistema de riego, luminosidad, lluvia, etc. pudiendo ser de 890 a 1681 plantas por hectárea, a una distancia de siembra entre 3.35x3.35 mts. Y 2.44x2.44 mts. en su orden.

1.7.5.2. SELECCIÓN DE SEMILLA.- Se procede a seleccionarla debiendo utilizarse aquellas cepas o semillas que provienen de “semilleros de plantaciones sanas”.

Existen tres tipos de cepas o material de propagación:

- a. Cepas de plantas maduras
- b. Cepas de plantas no maduras (esta es la mejor para plantarla)
- c. Cepas de hijos de espada.

1.7.5.3. PREPARACIÓN DE SEMILLA.- Las cepas deben sanearse y desinfectarse. El sanea se lo realiza con un machete pequeño (rabón) eliminando todas las raíces viejas y jóvenes, todos los tejidos viejos, las manchas negras causadas por enfermedades o pudrición, lesiones causadas por insectos hasta obtener un aspecto blanco limpio, este trabajo se realiza con cuidado para no lastimar las yemas de crecimiento, luego se lo sumerge en agua hirviendo durante 10 segundos como máximo.

Posteriormente se desinfecta la cepa con una solución que contenga insecticida o fungicida.

1.7.5.4. ALINEACIÓN Y ESTAQUILLADO.- Se procede a señalar en el terreno o sitio en el cual se va a realizar la siembra de la “cepa o semilla”, de acuerdo al sistema elegido, pudiendo ser en cuadro o en triángulo.

1.7.5.5. FORMA DE SIEMBRA.- Se procede a realizar el hoyado cavando el hoyo en cada señal para luego depositar las cepas a un lado. Los hoyos se deben de realizar en dimensiones normales que permitan a la cepa entrar sin dificultad.

Se lo coloca la cepa en sentido normal crecimiento, luego se le pone suelo suelto a su alrededor, aprisionándole para eliminar el aire del interior. Después de 6 a 8 semanas se procede a resembrar.

1.7.6. PRACTICAS AGRONÓMICAS

1.7.6.1. RIEGO.- El riego puede ser aplicado por gravedad, aspersión o inundación, dependiendo del sistema a emplearse, de la cantidad de agua disponible, tipo de suelo, topografía del mismo, disponibilidad económica y fertilidad del suelo.

La cantidad o frecuencia de riego depende de la calidad de agua, tipo de suelo, necesidades de cultivo, sistema utilizado y naturalmente la cantidad y distribución de las lluvias.

1.7.6.2. CONTROL DE MALEZAS.- El control de maleza se realiza en forma manual y en forma química mediante la aplicación de herbicidas o matamalezas. En el primer caso de control manual se realiza mediante “rozados o chapas” con machete, este control es eficaz pero no elimina definitivamente las malezas.

En el segundo caso para efectuar un control de malezas con productos químicos se debe tener conocimiento de las especies de maleza existentes para escoger el herbicida más adecuado.

1.7.6.3. FERTILIZACIÓN.- En los cultivos de banano del Ecuador se ha llegado a determinar que los elementos minerales indispensables y que deben ser aplicados al suelo son el nitrógeno y el potasio.

La fertilización debe ser adecuada y de acuerdo a las diferentes zonas o regiones variando la cantidad de fertilizantes a aplicarse.

El fertilizante debe ser aplicado en la zona de máxima absorción, es decir, más o menos desde la base de la planta hasta 1 mt., hacia fuera en semicírculo y alrededor del hijo seleccionado para producción.

Para fortalecer una fertilización racional y completa, que es lo que exige el banano por las características tan especiales de crecimiento, las cantidades de fertilizantes distribuidas en 4 aplicaciones por año, tomando en cuenta también la disposición del riego y el número de labores del cultivo, son las más recomendables para un buen aprovechamiento del producto por parte de la planta.

En cultivos de banano que no poseen riego se realiza 2 o 3 aplicaciones por año.

1.7.6.4. DESHIJE.- El deshije es una práctica cultural que tiene por objeto mantener la densidad adecuada por unidad de superficie, un espaciamento uniforme entre planta y planta, regular el número de hijos por unidad de producción, seleccionar los mejores hijos y eliminar los deficientes y excedentes.

Con deshije constante y eficiente se obtiene mayor producción y distribuida ésta durante todo el año, la unidad de producción, está constituida por la planta madre, el hijo y el nieto.

En una planta de banano hay tres clases de hijos:

- a. **Los hijos de espada o puyones:** son los que nacen profundos y alejados de la base de la planta madre, crecen fuertes y vigorosos. El follaje termina en punta, de ahí su nombre y es el mejor ubicado.
- b. **Los hijos de agua:** son los que desarrollan hojas anchas a muy temprana edad debido a deficiencias nutricionales. Siempre deben ser eliminados y se utilizan cuando hay un solo hijo de espada.
- c. **Los rebrototes:** son los hijos que vuelven a brotar luego de haber sido cortados, también desarrollan hojas anchas prematuramente y se parecen a los hijos de agua con la diferencia que se nota la cicatriz donde se hizo el corte. La rapidez de crecimiento de estos rebrototes decide la frecuencia de los deshijes.

Cuando se realiza el deshije los cortes con machetes deben hacerse lo más profundo posible tratando de eliminar la yema de crecimiento del hijo evitando el rebrote; el corte se dirige de adentro hacia fuera para no herir a la madre, luego se procede a cubrir la parte cortada.

1.7.6.5. DESHOJE.- Consiste en eliminar las hojas que ya cumplieron su ciclo u están agobiadas o dobladas y las que están interfiriendo el desarrollo del racimo. El corte debe ser lo más cercano posible a la base de la hoja; si una parte de una hoja joven y sana interfiere un racimo puede eliminarse esa parte rasgándola o cortándola, dejando el resto para que cumpla su función.

Esta labor debe ser constante según la frecuencia de la pérdida de hojas por parte de la planta.

1.7.6.6. APUNTALADO.- Es necesario realizar esta labor en toda la planta con racimo para evitar la caída y se ocasione pérdida de la fruta. Dentro de los materiales que sirven para este trabajo son: caña bambú, caña bravo, pambil, alambre, piola de yute, piola de plástico o nylon.

Cada productor escogerá el material que más le convenga a sus intereses; el más generalizado es la caña de bambú y caña brava, utilizando dos pedazos llamados palancas o cujes según la variedad cultivada y colocada en forma de tijera con el vértice hacia arriba y en tal posición que no topen el racimo.

1.7.6.7. ENFUNDE.- Es otra práctica que produce grandes beneficios al productor, consiste en proteger el racimo con una funda de polietileno perforada de dimensiones convenientes.

Se ha llegado a comprobar que la fruta enfundada tiene un 10% más de peso; por otra parte, la fruta está libre de la incidencia de daños causados por los insectos, por las hojas, por los productos químicos, por lo que se presenta limpia y excelente calidad.

La época más oportuna para el enfunde es cuando ha caído la tercera bráctea de la inflorescencia y queda abierta la correspondiente mano; se sujeta la funda al tallo de la inflorescencia a una altura conveniente. Se usa para sujetar: “sapan”, “piola de cabuya”, piola de algodón” o “cinta plástica de colores”, según la fecha para la cosecha por edad.

1.7.6.8. DESMANE.- Se procede a eliminar ocasionalmente la última mano o falsa mano y una o las dos siguientes que se estima no llegarán a adquirir el tamaño mínimo requerido, favoreciendo al desarrollo de las manos restantes.

El desmane se realiza cuando los frutos están colocados en dirección hacia abajo, sin usar herramienta alguna, solamente con la mano.

1.8. ENFERMEDADES DEL BANANO

Clasificación de las enfermedades

La clasificación de las enfermedades, de acuerdo a la naturaleza de los organismos patógenos son:

Virales

Bacterianas

Fúngicas

Producidas por algas

Danos por nematodos o picudos

Danos por exceso de sales en el suelo

Virales

- Clorosis infecciosa-cucumovirus
- Rayado necrótico-badnavirus

Bacterianas

- Tereque o moko (*Ralstonia solanacearum*)
- Pudrición acuosa del pseudotallo (*Erwinia carotovora* y *E. Chrysanthemi*)
- Pudrición del corno y cogollo (*Pseudomonas* sp. Y *Erwinia* sp.)

Fúngicas

- Sigatoka amarilla (*Mycosphaerella Musicota*)
- Sigatoka negra o raya negra (*Mycosphaerella Fijiensis*)

- Manchas del follaje (*Cordana Musae*)
- Mancha o rayado del envés (*Veronaea Musae*)
- Mancha irregular oscura de la hoja (*Deigtoniella Torulosa*)
- Salpicado de la hoja (*Periconiella Musae*)
- Mal de Panamá (*Fusarium oxysporum* fr. *Cubense*)
- Pudrición seca del tallo (*Marasmiellus Troyanus*)
- Enfermedades del fruto (*Verticillium Theobromae*)
 - ✓ *Colletotrichum Musae*
 - ✓ *Pyricularia Torulosa*
 - ✓ *Botryodiplodia Theobromae*
 - ✓ *Fusarium Roseum*

Producidas por algas

- *Cephaleuros Virescens*

Daños por nematodos o picudos

- *Radopholus similis*
- *Helicotylenchus multicinctus*
- *Meloidogyne* spp
- *Pratylenchus* spp
- *Rotylenchus* sp

Daños por exceso de sales en el suelo

- Carbonatos de calcio
- Cloruro
- Sulfato de sodio

CAPÍTULO II

LA SIGATOKA

INTRODUCCION

Las manchas de las hojas de los bananos y plátanos causadas por *Mycosphaerella fijiensis* Morelet, agente causal de la Sigatoka negra/raya negra y por *M. musicola* Leach ex Mulder, el agente causal de la Sigatoka amarilla, pueden considerarse desde el punto de vista económico las dos enfermedades más serias del género *Musa*.

Ambos patógenos atacan las hojas dando lugar a necrosis, a la disminución del tejido con actividad fotosintética y de los rendimientos brutos. Causan además la madurez prematura de los frutos lo cual representa pérdidas importantes del rendimiento neto exportable. En adición a esto, *M. fijiensis* ataca las hojas de clones de plátano que es un alimento básico en la dieta diaria en muchos países de América Latina, Asia y África. La distribución amplia de Sigatoka negra en África Occidental y en los trópicos de América la han convertido en la más costosa y dañina enfermedad de los bananos y plátanos (Stover y Simmonds, 1987). Por otro lado, la patogenicidad de *M. fijiensis* sobre clones de plátano resistentes a *M. musicola* a tenido un fuerte impacto social entre los pequeños agricultores debido a las infraestructuras de producción (pequeñas parcelas en poblados, cultivos inter cosechas) y el costo de los agroquímicos utilizados para el control.

Los costos de protección en las áreas donde la Sigatoka negra está presente han aumentado en más de cuatro veces en relación con la de Sigatoka amarilla (Presa, 1980). Stover y Simmonds (1987), han estimado que el costo de la protección contra Sigatoka negra en las grandes compañías de Centro América, superaba en 1985 el 27% de todos los costos de producción de una caja de banano. En la actualidad, con los problemas de pérdida de sensibilidad a los fungicidas triazoles, se han llegado a realizar hasta 40 tratamientos por año para lograr un adecuado nivel de control del patógeno y de calidad de la fruta (Pérez, 1996), lo que ha encarecido aun más la protección contra la enfermedad.

El primer informe de la Sigatoka procede de Java en 1902 y posteriormente en 1912, en la isla de Viti Levu, en Fiji (Meredith, 1970; Wardlaw, 1972). Se informaron pérdidas severas en Australia en 1924 y entre 1933 y 1934 su presencia fue informada en numerosas áreas del Caribe (Wardlaw, 1970; Stover, 1972). En 1937, había informes de daños serios en México, Centro América, Colombia, las Guyanas y las islas del Caribe (Stover y Simmonds, 1987). En Cuba, Popenoe informó su presencia en 1938 cuando el país exportaba 6 millones de racimos anuales y para 1940, la industria estaba totalmente colapsada por los daños de la enfermedad (Pérez, 1980). La última área importante de cultivo en ser invadida por la Sigatoka fue Ecuador, donde se dispersó rápidamente a partir de 1950 (Meredith 1970).

En 1963, se informó por primera vez la presencia de un tipo más severo de manchas de las hojas en Fiji (Rhodes, 1964), que fue denominada raya negra (**black leaf streak, maladie de raies noires**), debido al color pardo oscuro o negro de las rayas y manchas. Esta enfermedad causada por *Mycosphaerella fijiensis*, estaba distribuida en las islas del Pacífico mucho tiempo antes de su descubrimiento en Fiji en 1963. Según Stover (1978), los patrones de detección y distribución indican un centro de origen de *M. fijiensis* en el área de Nueva Guinea – Islas Salomón, desde donde el patógeno se movió a Taiwán, algunas islas adyacentes a Indonesia y a la mayoría de las islas del Pacífico sur, en hojas de espada en rizomas para plantación o en restos de hojas. En los estudios realizados de patrones de restricción del ADN (RFLP), con colecciones del hongo de diferentes partes del mundo, se determinó la mayor diversidad genética en las poblaciones del Sudeste Asiático incluido Papua-Nueva Guinea (Carlier et al. 1993). El patógeno según Stover se movió a Filipinas, Fiji, Hawai y el sur de Malasia en los últimos 35 años. En las áreas donde la Sigatoka negra se introduce, sustituye a la Sigatoka en un lapso de tres años aproximadamente.

La primera información de la presencia de *M. fijiensis* fuera de Asia, fue en Honduras en 1972 y en Zamba en 1973 (Stover, 1980 b). El patógeno fue encontrado en Cabón en 1977 y se dispersó a los países vecinos (el Congo, Guinea Ecuatorial, Camerún), estando presente en la actualidad en todos los

países del occidente de África. La Sigatoka negra se dispersó entre 1977 y 1980 al sur de México y todos los países de América Central; fue informada su presencia en Colombia en 1981, Ecuador, Venezuela en 1993, y en Perú en 1994. De las islas del Caribe, sólo Cuba en 1991 y Jamaica en 1995 han informado la presencia de la enfermedad (Pollard1995).

2.2. ANTECEDENTES

Tanto la Sigatoka amarilla causada por *M. musicola* como la Sigatoka negra causada por *M. fijiensis* son enfermedades específicas de las hojas de bananos y plátanos. Los daños a los frutos son indirectos y debido a los efectos de estas enfermedades sobre el área foliar. Cuando los daños son fuertes, los racimos tienden a ser más cortos y de menor peso y la pulpa de los frutos presenta un color amarillo con una fuerte tendencia a la madurez prematura.

La determinación de la velocidad de evolución de la enfermedad es de vital importancia para la instrumentación de las medidas de control en ambos casos y más específicamente para la determinación de las fechas de tratamiento. Para este fin, se requiere poder conocer las diferentes etapas por que transcurren los síntomas desde su aparición hasta su madurez y

senescencia. Brun (1958), describió los diferentes estados en la evolución de las manchas de Sigatoka de la siguiente forma:

Estado 1. Rayas de color verde muy pálido, menores a 1 mm de longitud, paralelas a las nervaduras de las hojas, visibles por el haz de las hojas a trasluz. Frecuentemente pasan inadvertidas si no se conocen. Este estado se corresponde con la penetración de las hifas de infección a la cámara subestomática y a la muerte de las células oclusivas y anexas de los estomas de las hojas.

Estado 2. Las estrías se van desarrollando y alcanzan la forma de una raya de varios milímetros de longitud, de color verde claro. El micelio que penetró por los estomas, se hace superficial al salir de nuevo por los mismos y aparece por las dos caras de la hoja, en forma de hifas carmelitosas de 3 – 4 μ de diámetro, las cuales crecen epifílicamente y vuelven a penetrar por otros estomas situados a 2 ó 3 mm de la mancha inicial. A este crecimiento, le corresponde un aumento de los daños quedando las hifas situadas a nivel de la lámina media de las paredes celulares y se necrosa el tejido en empalizada.

Estado 3. Las rayas comienzan a ensancharse, al mismo tiempo que aumentan de longitud. Los bordes no están bien definidos y se confunden

con el color normal de la hoja; el color comienza a cambiar a pardo rojizo. Los estromas comienzan a formarse en la cámara subestomática en el centro de la lesión.

Estado 4. La mancha toma una forma elíptica y alargada, definida, de color pardo oscuro; el centro de las lesiones se deprime progresivamente y se puede apreciar en su borde externo, la presencia de un halo amarillento brillante. Cuando la humedad es alta, se puede observar un halo de aspecto húmedo. En esta etapa, se diferencian los esporodoquios y si la humedad es favorable, comienza la producción de conidios.

Estado 5. Las manchas son ovales, más o menos alargadas; miden hasta 20 mm de longitud por 2 mm de ancho. El centro es grisáceo, deprimido, con el margen pardo oscuro a casi negro. Por fuera de este margen, puede observarse un halo amarillo. En esta etapa, cesa la producción de conidios y los peritecios y espermogonios del patógeno están presentes.

La caracterización de la evolución de los síntomas de Sigatoka negra ha sido realizada por diferentes autores (Rhodes, 1964, con material de Fiji; Meredith y Lawrence, 1968 con material de Hawaii; Stover y Dickson, 1975 en Honduras; Mulder y Stover, 1976 con material de diferentes procedencias; Fouré, 1982 en Gabón).

Siguiendo los criterios de Fouré (1982), las características de los síntomas son:

2.2.1. Estado 1

Aparece unas pequeñas puntuaciones (pecas) de color amarillo pálido, de un diámetro aproximado de 0.2 milímetro de forma irregular y difusas, solo perceptibles en el envés de las hojas. Esta puntuación amarilla no es siempre visible y pasa usualmente inadvertida. No es tampoco claramente visible en muchos clones. Cuando las condiciones son muy favorables al desarrollo de la enfermedad, estas pueden aparecer en la hoja 2 (contando la última hoja abierta como la hoja 1). Casi siempre aparecen en las hojas 3 y 4. estas pecas se alargan y alcanzan aproximadamente 1 milímetro de longitud, tomando las características de una raya pardo rojiza y no son visibles por el haz.

2.2.2. Estado 2

La raya se alarga hasta tomar una longitud variable. La principal características de este estado es que ya las rayas son visibles por el haz de las hojas. Siguen manteniendo su color pardo rojizo - café.

2.2.3. Estado 3

Las Rayas se alargan hasta alcanzar 20 - 25 mm. de longitud y 2 mm. de ancho, manteniendo su color café característico. Si la densidad de inóculo es muy grande, las rayas pueden coalescer formando zonas necróticas, que le dan un aspecto oscuro a las hojas. La distribución de estas rayas es variable; en ocasiones es muy numerosa en el semilimbo izquierdo de las hojas, pero en otras aparecen igualmente distribuidas en ambos. Frecuentemente, se agrupan en ambos lados del nervio central de la hoja y a veces varias rayas coalescen para formar rayas compuestas más grandes.

2.2.4. Estado 4

Puede considerarse el primer estado de mancha. Las rayas se ensanchan y toman un contorno más o menos redondeado, elíptico o fusiforme. La transición de rayas a manchas es caracterizada por el desarrollo de un borde acuoso o pardo claro alrededor de la mancha. Este borde acuoso se hace claramente visible temprano en la mañana cuando el rocío está presente o después de una lluvia.

2.2.5. Estado 5

El color pardo rojizo se torna pardo oscuro o casi negro, el área central que rodea la mancha se vuelve más pronunciada debido al oscurecimiento. En este estado, puede ocurrir un amarillamiento ligero del tejido de la hoja que rodea el borde acuoso de la mancha. Este estado caracteriza el color oscuro

casi negro que toma el follaje de las plantas afectadas seriamente por la enfermedad.

2.2.6. Estado 6

El centro de la mancha se seca, volviéndose gris claro y después se deprime, la mancha es rodeada de un borde estrecho bien definido pardo oscuro o negro. Entre este borde y el verde normal de la hoja hay un halo amarillento brillante en la zona de transición. Después que las hojas se han secado y colapsado, las manchas permanecen claramente visibles debido al centro claro y el borde oscuro.

No todas las manchas siguen esta secuencia de desarrollo, algunas no se desarrollan más allá del segundo - tercer estado. De forma igual a Sigatoka, en las plantas jóvenes las rayas iniciales se desarrollan a manchas ovales pardo – rojizas y las características de las primeras etapas de las rayas son de poca duración. En el caso de alta presión de inóculo aparecen sobre las hojas una gran cantidad de lesiones por unidad de área, generalmente de poco tamaño lo que provoca que a partir del estado 3, se necrosen los fragmentos provocando la aparición de un gran número de pseudotecios del patógeno.

2.3. MEDIDAS DE EXCLUSIÓN

La entrada de Sigatoka negra a una zona libre de la enfermedad o a plantaciones nuevas, debe ser dilatada tanto como sea posible. Aunque la enfermedad se puede trasladarse a través de las ascósporas transmitidas por el viento, desde focos de la enfermedad a zonas relativamente distantes situadas a favor de las corrientes de aire predominantes o por el paso de tormentas o huracanes, bajo condiciones normales, el movimiento de la enfermedad a zonas libre puede ser dilatado adoptando regulaciones cuarentenarias, especialmente entre zonas que de alguna forma están geofiguradamente aisladas. Entre las medidas a adoptar se encuentran:

- Debe limitarse el movimiento de material agámico de plantación procedente de plantaciones afectadas; la Sigatoka negra puede estar presente en las hojas lanceoladas de los hijos de espada.
- Preferiblemente usar vitroplantas procedentes de cultivo de tejidos que además van a estar libres de nemátodos, picudo negro y bacterias el género *Erwinia*. De no ser esto factible, procurar no mover residuos de hojas que puedan mover el hongo. La semilla agámica debe ser tratada con un fungicida previo a su plantación. Los territorios libres de Sigatoka negra deben organizar y promover programas de semilla

locales para los agricultores de forma que no sea necesario el movimiento desde zonas afectadas.

- Establecer regulaciones para el movimiento libre de fruta, sin beneficiar y envasar con las que se pueda por diferentes motivos trasladar hojas o residuos de hojas desde zonas afectadas a libres.

La lucha contra la Sigatoka negra y amarilla es esencialmente la lucha contra el inóculo de ambos patógenos. Para poder lograr un adecuado control de la enfermedad es esencial eliminar todas las formas de inóculo disponibles y generalizar el control al nivel de una zona o territorio de una forma tan generalizada como sea posible. Siendo *Mycosphaerella spp.* patógenos específicos del género *Musa* las fuentes de inóculos son esencialmente plantas de bananos o plátanos susceptibles que no reciben una adecuada protección. Las fuentes de inóculo más comunes están constituidas por:

- Parcelas de pequeños agricultores que no reciben protección, que poseen clones susceptibles. En el caso de Sigatoka negra, son generalmente plátanos o algunos clones Cavendish, que no reciben ningún nivel de protección contra la enfermedad. Una solución para esto, es la capacitación de los agricultores de la región, de las características de ambas enfermedades y divulgación de las cualidades de los nuevos clones con resistencia

a Sigatoka negra y amarilla, junto a la implementación de un programa de distribución de materiales de plantación resistentes así como la promulgación donde sea factible, de medidas regulatorias de la plantación de material susceptible, especialmente en los casos de colindancia de grandes explotaciones compactas de clones Cavendish y plátanos altamente susceptibles a la enfermedad.

- Plantaciones y rebrotes abandonados de clones susceptibles a la enfermedad. Estos rebrotes en las colindancias de las explotaciones comerciales deben ser totalmente destruidos o sometidos a tratamientos hasta su total destrucción.
- Focos calientes en el interior de las plantaciones debido a condiciones locales muy favorables a la enfermedad o a malas prácticas de manejo agronómico y de la protección. Entre estas, zonas de plantaciones mal protegidas durante los tratamientos aéreos por una configuración irregular de los campos o la existencia de obstáculos; inadecuado solapamiento de las bandas de tratamiento por un deficiente estaquillado de los campos o trabajo de los bandereros; exceso de población e inadecuado deshoje y deshije; problemas de drenaje o microclima muy húmedo por cercanía a lagunas, ríos, presas, etc. todas las condiciones que determinan el surgimiento de focos calientes, deber ser eliminadas

o recibir una atención diferenciada. En las localidades donde hay obstáculos permanentes para la buena calidad de la protección, debe procederse invariablemente a la plantación con clones resistentes a la enfermedad.

2.4. EVALUACIÓN DE LA ENFERMEDAD

Se debe tener un programa de control de la enfermedad de la planta en cada finca, para así prevenir daños, realizando evaluaciones periódicas pueden ser éstas semanales, quincenales sobre la evolución y el grado de severidad de la Sigatoka Negra.

Tenemos dos grupos de variables que determinan la enfermedad de la planta:

2.4.1. PREAVISO

Para poder conocer la evolución de la planta debemos registrar información como son los síntomas de la enfermedad, para ello se toma en cuenta la Hoja II, Hoja III y Hoja IV, en donde se cuenta el número de pizcas y estrías presentes y se las califica tomando las siguientes características:

+1 si existen más de 50 pizcas que indican la presencia de la sigatoka negra.

-1 si existen menos de 50 pizcas que indican la presencia de la sigatoka negra.

+2 si existen más de 50 estrías que indican la presencia de la sigatoka negra.

-2 si existen menos de 50 estrías que indican la presencia de la sigatoka negra.

Los datos de las variables que analizaremos resultan de cálculos matemáticos realizados por el CIBE.

2.4.2. SEVERIDAD

La severidad consiste en examinar todas las hojas de la planta del banano y considerando el porcentaje del área afectada, la clasificación según los porcentajes de severidad presentamos en la siguiente figura, tomando valores entre 1 y 6.

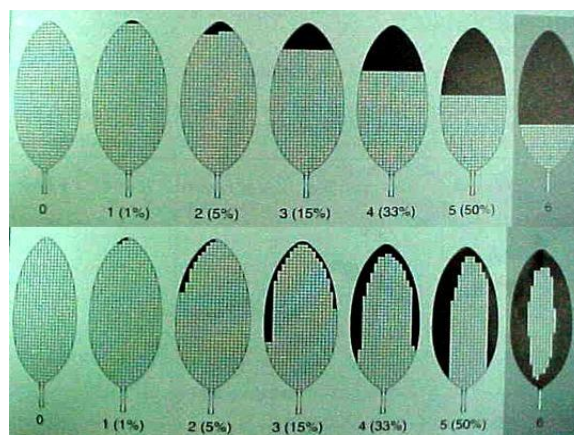


Fig. 2.1. Porcentaje de clasificación de la severidad

CAPÍTULO III

MÉTODOS ESTADÍSTICOS

INTRODUCCIÓN

A medida que aumenta la complejidad de nuestro mundo, se hace cada vez más difícil tomar decisiones inteligentes y bien documentadas. Con frecuencia tales decisiones deben tomarse con mucho menos que un conocimiento adecuado y experimentando una gran incertidumbre.

La estadística utiliza el diseño de experimentos y de muestreo para obtener una cantidad determinada de información a un costo mínimo y del uso óptimo de esta información para hacer inferencias con respecto a una población a partir de la información contenida en una muestra y proporcionar una probabilidad que permita realizar una inferencia acerca de alguna característica de la población de estudio. Hacer inferencias implica la descripción parcial o total de un fenómeno u objeto físico.

3.2. DEFINICIONES BÁSICAS.

3.2.1. POBLACIÓN

Es la colección completa de todas las unidades de observación de interés para el investigador.

3.2.2. PARÁMETRO

Es una medida descriptiva de la población total de todas las observaciones de interés para el investigador.

3.2.3. MUESTRA

Es una parte representativa de la población que se selecciona para ser estudiada ya que la población es demasiado grande como para analizarla en su totalidad.

3.2.4. ESTADÍSTICO

Es el elemento que describe una muestra y sirve como una estimación del parámetro de la población correspondiente.

3.2.5. VARIABLE

Es una característica de la población que se está analizando en un sentido estadístico.

Una variable puede ser cuantitativa o cualitativa. Si las observaciones pueden expresarse numéricamente, entonces es una *variable cuantitativa*.

Una *variable cualitativa* se mide de manera no numérica.

Además las variables cuantitativas pueden ser continuas o discretas.

Una *variable continua* es aquella que puede tomar cualquier valor dentro de un rango dado. No importa que tan cerca puedan estar dos observaciones, si el instrumento de medida es lo suficientemente preciso, puede hallarse una tercera observación que se encuentre entre las dos primeras. Una variable continua generalmente resulta de la medición.

Una *variable discreta* está limitada a ciertos valores, generalmente números enteros. Con frecuencia son el resultado de la enumeración o del conteo.

3.2.6. MUESTREO

Gran parte del trabajo de un estadístico se realiza con muestras. Las muestras son necesarias debido a que con frecuencia las poblaciones son demasiadas grandes para ser estudiadas en su totalidad. Es muy costoso y demanda mucho tiempo examinar la población total, por tanto, debe seleccionarse una muestra de la población, calcular el estadístico de la muestra, y utilizarlo para estimar el parámetro correspondiente de la población.

Este análisis sobre las muestras implica una distinción entre las dos principales ramas del análisis estadístico.

1. **Estadística Descriptiva**, es el proceso de recolectar, agrupar y presentar datos de una manera tal que describa fácil y rápidamente dichos datos.
2. **Estadística Inferencial**, involucra la utilización de una muestra para sacar alguna inferencia o conclusión sobre la población de la cual hace parte la muestra.

La exactitud de toda estimación es de enorme importancia. Esta exactitud depende, en gran parte, de la forma cómo se toma la muestra, y del cuidado que se tenga para garantizar que la muestra proporcione una imagen confiable de la población. Sin embargo, con mucha frecuencia se comprueba que la muestra no es del todo representativa de la población y resultará un error de muestreo.

3.2.7. ERROR DE MUESTREO

Es la diferencia entre el parámetro desconocido de la población y el estadístico de la muestra utilizado para calcular el parámetro.

Existen dos causas posibles del error muestreo. Debido al factor azar en la selección de elemento de la muestra, es posible seleccionar sin darse cuenta, elementos atípicos que no representan la población. Por otro lado, el azar puede producir un gran número de elementos de muestra que sean

inusualmente pequeños, produciendo una subestimación del parámetro. En cualquiera de los dos casos, ha ocurrido un error de muestro.

3.2.8. SESGO MUESTRAL

Es la tendencia a favorecer la selección de ciertos elementos de muestra en lugar de otros.

3.2.9. MEDIDAS DE LA TENDENCIA CENTRAL

Existen tres métodos comunes para identificar el centro de un conjunto de datos: la media, la mediana y la moda.

3.2.9.1. LA MEDIA o media aritmética, es la medida de la tendencia central que usualmente se le llamaba promedio.

3.2.9.1.1. LA MEDIA DE UNA POBLACIÓN es el parámetro μ (que se pronuncia miu). Si hay N observaciones en el conjunto de datos de la población, la media se calcula así:

$$\text{Media poblacional} \quad \mu = \frac{X_1 + X_2 + X_3 + \dots + X_N}{N}$$

$$\mu = \frac{\sum_{i=1}^N X_i}{N}$$

La letra griega mayúscula Σ es el signo de sumatoria que indica que se suman todas las observaciones de 1 a N. Las x_i denotan las observaciones individuales. En aras de la simplicidad, el subíndice y el superíndice se reducirán y aparecerá el signo de la sumatoria como simplemente Σ .

3.2.9.1.2. LA MEDIA DE UNA MUESTRA, es un estadístico \bar{x} (que se lee x barra). Con n observaciones en el conjunto de datos de la muestra, la media se determina así:

$$\text{Media muestral} : \bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

3.2.9.2. LA MEDIANA

Algunas veces es llamada media posicional, porque queda exactamente en la mitad del conjunto de datos después de que las observaciones se han colocado en serie ordenada. La mitad (50%) de las observaciones estará por encima de la mediana, la otra mitad estará por debajo de ella.

Si el conjunto de datos tiene un número impar de observaciones, la posición

de la mediana es: $\frac{n+1}{2}$.

3.2.10. MEDIDAS DE DISPERSIÓN

Una descripción más completa del conjunto de datos puede obtenerse si se mide que tan dispersos están los datos alrededor de dicho punto central. Esto es precisamente lo que hacen las medidas de dispersión. Indican cuánto se desvían las observaciones alrededor de su media.

3.2.10.1. EL RANGO

La medida de dispersión más simple (y menos útil) es el rango o recorrido. El rango es simplemente la diferencia entre la observación más alta y la más baja. Su ventaja es que es fácil de calcular. Su desventaja es que considera solo dos de los cientos de observaciones que hay en un conjunto de datos. El resto de las observaciones se ignoran.

3.2.10.2. VARIANZA Y DESVIACIÓN ESTÁNDAR DE UNA POBLACIÓN

La varianza y la desviación estándar son medidas de dispersión mucha más útiles. Proporcionan una medida más significativa sobre el punto hasta el cual se dispersan las observaciones alrededor de su media.

Varianza es el promedio de las observaciones respecto a su media elevadas al cuadrado. La varianza para una población σ^2 (se lee sigma al cuadrado) es:

$$\text{varianza poblacional } l : \sigma^2 = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 + \dots + (x_N - \mu)^2}{N}$$

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

En donde $x_1, x_2, x_3, \dots, x_N$ son las observaciones individuales.

μ es la media poblacional

N es el número de observaciones

La desviación estándar σ es:

$$\text{Desviación estándar poblacional } l : \sigma = \sqrt{\sigma^2}$$

3.2.10.3. VARIANZA O DESVIACIÓN ESTÁNDAR PARA UNA MUESTRA

La varianza y la desviación estándar para una muestra representan medidas de dispersión alrededor de la media. Se calculan de manera parecida a aquellas para una población. La varianza de la muestra s^2 es:

varianza de muestra :

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Y la desviación estándar de la muestra es:

Desviación estándar :

$$s = \sqrt{s^2}$$

de la muestra

3.2.10.4. OTRAS MEDIDAS DE DISPERSIÓN

Aunque la varianza y la desviación estándar son las medidas de dispersión más útiles en análisis estadístico, existen otras medidas con las cuales puede describirse la dispersión de un conjunto de datos. Estas medidas adicionales de dispersión son los *cuartiles*, *deciles* y *percentiles*.

3.2.10.4.1. CUARTILES, DECILES Y PERCENTILES

CUARTILES, cada conjunto de datos tiene tres cuartiles que lo dividen en cuatro partes iguales. El primer cuartil es ese valor debajo del cual clasifica el 25% de las observaciones, y sobre el cual puede encontrarse el 75% restante. El segundo cuartil representa el valor de la variable aleatoria acumula al menos el 50% de los datos, esto significa que se puede tener la

mitad de las observaciones por debajo y la otra mitad por encima de ese valor; por lo tanto, es lo mismo que la mediana. El tercer cuartil es el valor debajo del cual está el 75% de las observaciones y encima del cual puede encontrarse el 25% restante.

Los **DECILES** separan un conjunto de datos en 10 subconjuntos iguales, y los **PERCENTILES** en 100 partes. El primer decil es la observación debajo de la cual se encuentra el 10% de las observaciones, mientras que el 90% restante se encuentra encima de éste. El primer percentil es el valor debajo del cual se encuentra el 1% de las observaciones, y el resto está encima de éste. Puede aplicarse una interpretación similar al resto de deciles y percentiles. Todo conjunto de datos tiene 9 deciles y 99 percentiles.

Un percentil y su ubicación en un arreglo ordenado se identifican mediante los subíndices. Por ejemplo, el decimoquinto percentil se indica como P_{15} , y su ubicación en la serie ordenada es L_{15} .

$$\text{Ubicación de un percentil} : L_p = (n + 1) \frac{P}{100}$$

En donde:

L_p es el sitio del percentil deseado en una serie ordenada.

n es el número de observaciones

P es el percentil deseado

Regresando a los deciles y cuartiles, se nota que el primer decil es igual a P_{10} , el segundo decil es igual a P_{20} , y así sucesivamente. Adicionalmente, el primer cuartil es igual a P_{25} , el segundo cuartil es igual a P_{50} , y P_{75} es el tercer cuartil.

3.2.10.4.2. RANGO O RECORRIDO INTERCUÁRTICO (RIQ)

El RIQ es la diferencia entre el tercer cuartil y el primer cuartil. Es decir, $P_{75} - P_{25}$. La mitad de las observaciones se clasifican dentro de éste rango. Consta del 50% de la mitad de las observaciones y corta el 25% inferior y el 25% superior de los puntos de datos. Como resultado, el RIQ proporciona una medida de dispersión que no está muy influenciada por unas cuantas observaciones extremas.

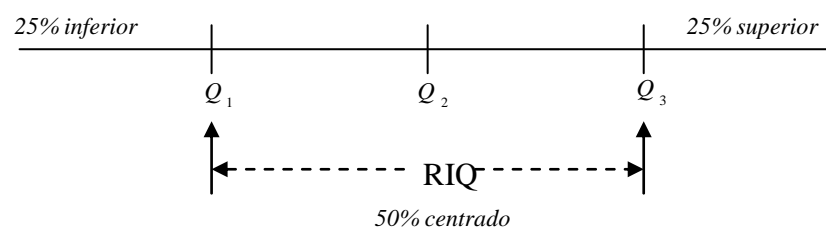


Fig. 3.1 RIQ

3.3. USO FRECUENTES DE LA DESVIACIÓN ESTÁNDAR

3.3.1. LA DISTRIBUCIÓN NORMAL

Una distribución normal es una distribución de datos continuos (no discretos) que produce una curva simétrica en forma de campana.

En una distribución normal, la media, la mediana y la moda son todas iguales. Es de importancia que la mitad del área que está bajo la curva está a la izquierda de la media y la otra mitad del área que está por debajo de la curva está a la derecha de la media.

3.3.2. LA REGLA EMPÍRICA

Si se incluyen todas las observaciones que están a una desviación estándar de la media (una desviación estándar por encima de la media y una desviación estándar por debajo de la media) éstas serán el 68,3% de todas las observaciones. Es decir, que no importa cual es la media ni cual es la desviación estándar, se puede estar seguro que el 68,3% de las observaciones quedan a una desviación estándar de la media si las observaciones están distribuidas normalmente.

3.3.3. SESGO

No todas las distribuciones son normales. Algunas están sesgadas a la izquierda o a la derecha.

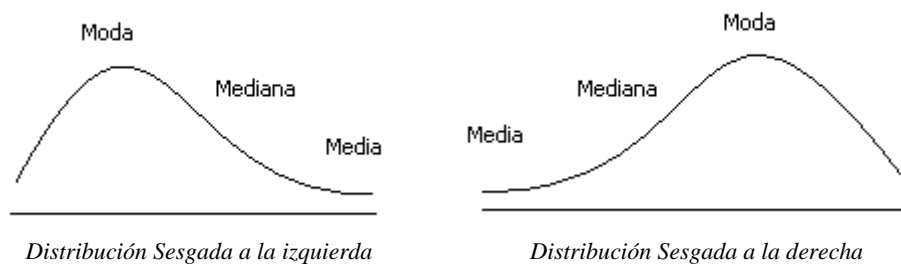


Fig. 3.2 Curvas Sesgadas

En ambos casos, la moda es por definición la observación que ocurre con mayor frecuencia. Por tanto, está en el pico de la distribución. Sin embargo por su naturaleza, la media se ve más afectada por las observaciones extremas. Por tanto, es halada en la dirección del sesgo, más de lo que está la mediana, la cual está en algún sitio entre la media y la moda.

El sesgo puede medirse mediante el **Coefficiente de Sesgo de Pearson**.

$$\text{Coeficiente de sesgo} : P = \frac{3(\bar{x} - \text{mediana})}{s}$$

Si $P < 0$, los datos están sesgados a la izquierda, si $P > 0$, entonces están sesgadas a la derecha; si $P = 0$ están distribuidos normalmente.

3.4. MEDIDAS DE DISTRIBUCIÓN

Las medidas de distribución nos permiten identificar la forma en que se separan o aglomeran los valores de acuerdo a su representación figura. Estas medidas describen la manera como los datos tienden a reunirse de acuerdo con la frecuencia con que se hallen dentro de la información. Su utilidad radica en la posibilidad de identificar las características de la distribución sin necesidad de generar figura. Sus principales medidas son la *Asimetría* y la *Curtosis*.

3.4.1. ASIMETRÍA

Esta medida nos permite identificar si los datos se distribuyen de forma uniforme alrededor del punto central (Media Aritmética). La asimetría presente tres estados diferentes (véase Fig.3.3), cada uno de los cuales define de forma consiga como están distribuidos los datos respecto al eje de asimetría. Se dice que la *asimetría* es *positiva* cuando la mayoría de los datos se encuentran por encima del valor de la media aritmética, la curva es *simétrica* cuando se distribuyen aproximadamente la misma cantidad de valores en ambos lados de la media y se conoce como *asimetría negativa* cuando la mayor cantidad de datos se aglomeran en los valores menores que la media.

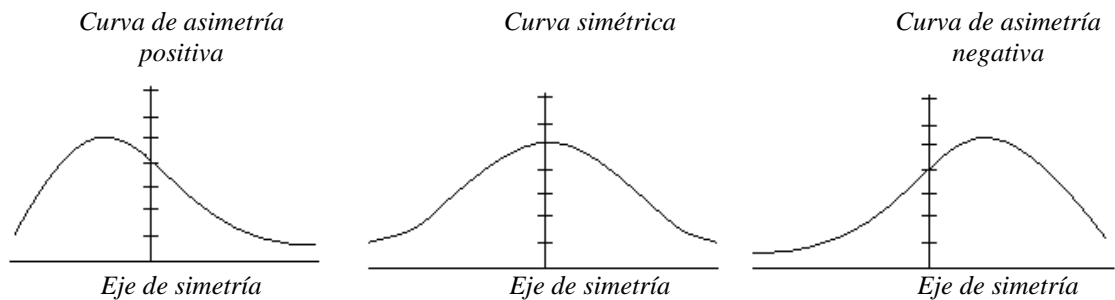


Fig. 3.3 Tipos de curvas asimétricas

El *coeficiente de asimetría*, se representa mediante la ecuación matemática:

$$g_1 = \frac{\frac{1}{n} \sum (x_i - \bar{x})^3 * n_i}{\left(\frac{1}{n} \sum (x_i - \bar{x})^2 * n_i \right)^{\frac{3}{2}}}$$

Donde g_1 representa el coeficiente de asimetría de Fisher, x_i cada uno de los valores, \bar{x} la media de la muestra y n_i la frecuencia de cada valor. Los resultados de ésta ecuación se interpretan:

- $g_1 = 0$: se acepta que la distribución es simétrica, es decir, existe aproximadamente la misma cantidad de valores a los dos lados de la media. Este valor es difícil de conseguir por lo que se tiende a tomar los valores que son cercanos ya sean positivos o negativos (± 0.5)
- $g_1 > 0$: la curva es asimétricamente positiva por lo que los valores se tienden a reunir más en la parte izquierda que en la derecha de la media.

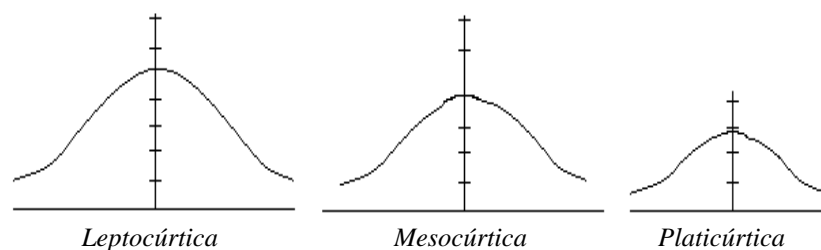
- $g_1 < 0$: la curva es asimétricamente negativa por lo que los valores se tienden a reunir más en la parte derecha de la media.

Desde luego entre mayor sea el número (positivo o negativo), mayor será la distancia que separa la aglomeración de los valores con respecto a la media.

3.4.2. CURTOSIS

Esta medida determina el grado de concentración que presentan los valores en la región central de la distribución (vea Fig.3.4). Por medio del *coeficiente de curtosis*, podemos identificar si existe una gran concentración de valores (*Leptocúrtica*), una concentración normal (*Mesocúrtica*) ó una baja concentración (*Platicúrtica*).

Fig.3.4 Tipos de curtosis



Para calcular el coeficiente de curtosis se utiliza la ecuación:

$$g_2 = \frac{\frac{1}{n} \sum (x_i - \bar{x})^4 * n_i}{\left(\frac{1}{n} \sum (x_i - \bar{x})^2 * n_i \right)^2} - 3$$

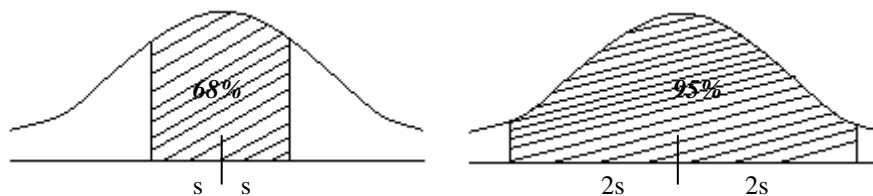
Donde g_2 representa el coeficiente de curtosis, x_i cada uno de los valores, \bar{x} la media de la muestra y n_i la frecuencia de cada valor. Los resultados de ésta ecuación se interpretan:

- $g_2 = 0$: la distribución es *Mesocúrtica*: Al igual que en la asimetría es bastante difícil encontrar un coeficiente de curtosis de cero (0), por lo que se suelen aceptar los valores cercanos (± 0.5).
- $g_2 > 0$: la distribución es leptocúrtica.
- $g_2 < 0$: la distribución es platicúrtica.

Cuando la distribución de los datos cuenta con un coeficiente de asimetría $g_1 = \pm 0.5$ y un coeficiente de curtosis de $g_2 = \pm 0.5$, se le denomina Curva Normal. Este criterio es de suma importancia ya que para la mayoría de los procedimientos de la estadística de inferencia se requiere que los datos se distribuyan normalmente.

La principal ventaja de la distribución normal radica en el supuesto que el 95% de los valores se encuentra dentro de una distancia de dos desviaciones estándar de la media aritmética (véase Fig. 3.5); es decir, si tomamos la media y le sumamos dos veces la desviación y después le restamos a la media dos desviaciones, el 95% de los casos se encontraría dentro del rango que compongan estos valores.

Fig. 3.5 Curva Normal



3.5. COEFICIENTE DE VARIACIÓN

Con frecuencia debemos considerar el *Coefficiente de Variación (CV)*, el cual sirve como medida relativa de dispersión. El coeficiente de variación determina el grado de dispersión de un conjunto de datos relativo a su media. Se calcula dividiendo la desviación estándar de una distribución por su media y multiplicando por 100.

$$\text{Coeficiente de variación : } CV = \frac{S}{x} (100 \%)$$

3.6. FIGURA DE CAJAS (BOX-PLOT)

El figura de cajas (*box-plot*), tiene por objeto reflejar mejor la forma de distribuciones dando en un mismo elemento figura información acerca de la mediana, la media, los cuartiles 5%, 25%, 75% y 95% y mostrando la presencia, si los hubiere, de valores extremos. La especificación de las variables en el selector de variables de este tipo de figura es idéntica que para el diagrama de puntos.

3.7. BIPLOTS

Las figuras de dispersión son usadas para visualizar directamente las observaciones o las variables, las relaciones en otra dimensión son sólo implícitas. Las figuras Biplots, muestran las observaciones y las variables en la misma figura, de forma tal que se pueden hacer interpretaciones sobre las relaciones conjuntas. El prefijo “bi” en el nombre biplot refleja la característica de que tanto observaciones como variables son representadas en el mismo figura.

En los Biplots, las observaciones son generalmente graficadas como puntos. La configuración de los puntos es obtenida a partir de combinaciones lineales de las variables originales. Las variables son graficadas como vectores

desde el origen. Los ángulos entre las variables representan la correlación entre las variables.

Las dimensiones seleccionadas para el Biplot son aquellas que mejor explican la variabilidad de los datos originales. Para encontrar los ejes óptimos para la graficación de observaciones y variables en un espacio común se utiliza la idea de que cualquier matriz de datos $n \times p$, puede ser representada aproximadamente en d -dimensiones como el producto de dos matrices, $A(n \times d)$ y $B(p \times d)$, y d es el rango de la matriz original, así AB' aproxima la matriz original. Debido a que A y B tienen una base común de d vectores, es posible mostrar las filas y las columnas de la matriz original sobre la misma figura con varias condiciones de optimalidad y con la posibilidad de realizar interpretaciones sobre las distancias entre puntos.

Los Biplots pueden considerarse como una técnica de reducción de dimensión ya que las filas de A representan las observaciones en un espacio de menor dimensión (puntos filas) y las columnas de B' representan las variables (puntos columnas) en ese mismo espacio. Dada la descomposición por valor singular de X como $X = UD_TV'$, donde U es $n \times p$ con columnas ortogonales, V es una matriz ortogonal $p \times p$ y D_T una matriz diagonal $p \times p$ de valores singulares, las matrices A y B pueden ser expresadas como:

$$A = UD_r^\alpha \text{ y } B = UD_r^{1-\alpha}$$

Donde α es usualmente igual a 0, $\frac{1}{2}$ ó 1 para proveer condiciones de optimalidad en el figura (Gower y Digby, 1981). Los Biplots son figuras de dispersión de los $n + p$ vectores de A y B en un mismo espacio d -dimensional. Comúnmente se presentan figuras bidimensionales seleccionando los dos componentes en U y V asociados con los dos valores singulares más altos en D . Estos figuras son aproximaciones de la matriz original al menos que toda la variabilidad sea explicada por los dos primeros ejes. Los Biplots pueden ser utilizados para representar los resultados del ACP, estos se obtienen realizando dos ACP, uno sobre la matriz de covarianzas de las observaciones y otro sobre la matriz de covarianzas de las variables y graficando ambos autovectores correspondientes a la variables convenientemente escalados en el mismo espacio en que se graficaron las observaciones.

En los Biplots la distancia entre símbolos representando observaciones y símbolos representando variables no tienen interpretación, pero las direcciones de los símbolos desde el origen si pueden ser interpretadas. Las observaciones (puntos filas) que se grafican en una misma dirección que una variables (punto columna) podría tener valores relativamente altos para esa variable y valores bajos en variables o puntos columnas que se grafican en dirección opuesta. Dependiendo de las condiciones de optimalidad

especificadas, las distancias entre los puntos filas o entre los puntos columnas pueden ser estadísticamente interpretadas, los ángulos entre los vectores que representan las variables pueden ser interpretados en términos de las correlaciones entre variables y las longitudes de los rayos pueden hacerse proporcionales a las desviaciones estándar. Cuando las longitudes de los vectores son similares la figura sugiere contribuciones similares de cada variable en la representación realizada.

3.8. IMPUTACIÓN DE DATOS

En el análisis de grandes volúmenes de información es muy frecuente encontrar la existencia de datos con valores perdidos. De hecho, los datos completos constituyen más una excepción que la regla.

Esta situación constituye una severa limitación, puesto que los métodos estadísticos tradicionales están diseñados para ser aplicados sobre conjuntos de datos completos. Los continuos avances de la informática han hecho posible el surgimiento y puesta de nuevas metodologías para el tratamiento de información con datos faltantes, los cuales, en su mayoría, producen resultados aceptables cuando hay pocos valores perdidos.

Alternativas de Solución

1. *Descartar los individuos o variables con valores perdidos.*

Esta modalidad resulta atractiva debido a su simplicidad, ya que es posible analizar los datos sin necesidad de modificar el método estadístico utilizado. Solamente es recomendable en situaciones con muy pocos valores perdidos, diseminados a su vez sobre muy pocas unidades, puesto que de otra manera la pérdida de información parcial útil puede llegar a ser considerable.

2. *Estimar los valores perdidos utilizando algún método de imputación*

Una segunda opción consiste en utilizar la información contenida en los datos observados, para construir estimaciones “razonables” de los valores perdidos. Al ser imputadas tales estimaciones en las observaciones perdidas, se obtiene un juego de datos completo, susceptible de ser analizado con los métodos estadísticos estándar.

Una de las estrategias usualmente utilizadas requiere una matriz de datos con propiedades conocidas, de la cual se elimina aleatoriamente una cierta proporción de valores asumidos como perdidos, y que son posteriormente estimados e imputados.

Luego se construyen índices, tanto de bondad de las estimaciones como del grado de fidelidad con que son reconstruidas las propiedades de la matriz original a partir de la matriz imputada.

3.8.1. IMPUTACIÓN POR MEDIAS

Este método, propuesto por primera vez por Wilks (1932), es posiblemente uno de los procedimientos de imputación más antiguo y más sencillo.

Consiste en estimar los valores perdidos de la j -ésima variable mediante la media de sus valores observados, la cual ha sido llamada por Little y Rubin (1987), media de los valores perdidos.

3.8.2. IMPUTACIÓN POR REGRESIÓN (POLINÓMICA)

Este método, propuesto por Buck (1960), supone que las filas de la matriz de datos constituyen una muestra aleatoria de una población normal multivariante. El vector de medias y la matriz de varianzas y covarianzas de los datos completos son utilizados como estimaciones de los parámetros poblacionales, con los cuales se ajustan ecuaciones en regresión para cada una de las variables con datos perdidos, en término de las restantes.

Objetivo: Explicar o predecir la variable Y a través de una covariable X , a través de un polinomio de grado p :

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_p X^p + \varepsilon$$

Al añadir potencias de X , aumentamos la flexibilidad del modelo.

Casos particulares:

- $p = 1 \rightarrow$ recta de regresión
- $p = 2 \rightarrow$ regresión cuadrática
- $p = 3 \rightarrow$ regresión cúbica

Si la muestra consta de n pares $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, entonces los estimadores **mínimos cuadrados** de $(\beta_0, \beta_1, \beta_2, \dots, \beta_p)$ se obtienen:

$$\hat{\beta} / \underset{\beta}{\text{Min}} \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n [(y_i - \hat{y}_i)^2] = \sum_{i=1}^n [(y_i - (\beta_0 + \beta_1 x_i + \dots + \beta_p x_i))^2]$$

3.9. BONDAD DE AJUSTE (KOLMOGOROV)

Kolmogorov Smirnov permite probar si la muestra disponible se ajusta a un modelo distribucional teórico. Se supone que se dispone de una muestra aleatoria y que se desea probar si la distribución empírica se ajusta a alguna de las siguientes distribuciones: Normal (m,v), T de Student (v), F de Snedecor (u,v), Chi Cuadrado (v), Gamma (r,lambda), Beta (a,b), Weibull (a,b), Exponencial (lambda) o Gumbel (a,b). La distribución teórica debe ser completamente especificada (parámetros conocidos). Las hipótesis son:

$$H_0 : G(x) = F_{teórica}(x)$$

vs

$$H_1 : G(x) \neq F_{teórica}(x) \text{ para al menos una } x$$

Donde $G(x)$ es la función de distribución empírica (o de los valores observados) y $F_{teórica}(x)$ es la función de distribución teórica especificada por el usuario. La prueba es sensible a cualquier discrepancia entre las distribuciones (dispersión, posición, simetría, etc.). El estadístico se basa en la máxima diferencia entre las dos distribuciones y se define como la máxima distancia vertical entre $G(x)$ y $F_{teórica}(x)$. El Estadístico de Kolmogorov es:

$$D = \sup_x \left\{ \left| F_{teórica}(x) - G(x) \right| \right\}$$

La prueba de bondad de ajuste de Kolmogorov se usa cuando la función de distribución hipotetizada se especifica completamente, es decir que no se necesita estimar ningún parámetro desconocido de la muestra. Se deberá reportar el valor del estadístico (D) y el valor p correspondiente obtenido desde la aproximación asintótica de la distribución del estadístico.

Valores de p menores al nivel de significación sugieren el rechazo de la H_0 .

3.10. COEFICIENTE DE CORRELACIÓN DE PEARSON

El *coeficiente de correlación de Pearson* es una medida de la magnitud de la asociación lineal entre dos variables que no depende de las unidades de medida de las variables originales. Para las variables j-ésima y k-ésima se define como:

$$r_{jk} = \frac{S_{jk}}{\sqrt{S_j^2 S_k^2}} = \frac{\left(\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \right) / (n-1)}{\sqrt{\left(\left(\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \right) / (n-1) \right) \left(\left(\sum_{i=1}^n (x_{ik} - \bar{x}_k)^2 \right) / (n-1) \right)}}$$

Donde S_{jk} es la covarianza entre la variable K y la variable k, S_j^2 y S_k^2 son las varianzas de las variables j y k respectivamente.

El coeficiente de correlación muestral representa la covarianza de los valores muestrales estandarizados. Asume valores en el intervalo [-1;1] y el signo indica la dirección de la asociación (valores negativos se producen cuando la tendencia promedio indica que si un valor en el par observado es más grande que su media, el otro valor es más pequeño que su media).

3.11. DISTRIBUCIÓN JI CUADRADO

Una distribución muestral que puede ser definida en términos de variables aleatorias normales es la distribución ji cuadrado o χ^2 . Si Z_1, Z_2, \dots, Z_k son variables aleatorias independientes y normalmente distribuidas con media 0 y varianza 1, lo cual se simboliza como NID(0,1), entonces la variable aleatoria

$$\chi_k^2 = Z_1^2 + Z_2^2 + \dots + Z_k^2$$

Sigue una distribución ji cuadrada con k grados de libertad. La función de densidad ji cuadrada es:

$$f(\chi^2) = \frac{1}{2^{k/2} \Gamma\left(\frac{k}{2}\right)} (\chi^2)^{(k/2)-1} e^{-\chi^2/2} \quad \chi^2 > 0$$

3.12. DISTRIBUCIÓN F

Si χ_u^2 y χ_v^2 son dos variables aleatorias independientes con distribuciones ji cuadrada con u y v grados de libertad, respectivamente, entonces la razón

$$F_{u,v} = \frac{\chi_u^2 / u}{\chi_v^2 / v}$$

tiene una distribución F con u grados de libertad en el

numerador y v grados de libertad en el denominador. La distribución de probabilidad de F es:

$$h(F) = \frac{\Gamma\left(\frac{u+v}{2}\right) \left(\frac{u}{v}\right)^{u/2} F^{(u/2)-1}}{\Gamma\left(\frac{u}{2}\right) \Gamma\left(\frac{v}{2}\right) \left[\left(\frac{u}{v}\right)F + 1\right]^{(u+v)/2}} \quad 0 < F < \infty$$

3.13. ANÁLISIS DE CONGLOMERADOS

Las representaciones que se estudian en este tema tienen su antecedente histórico en el sistema taxonómico de los seres vivos debido al naturalista Carl Von Linné.

Sea ε un conjunto de n objetos o individuos sobre los que se ha calculado alguna medida de distancia.

Sea $D = (\delta_{ij})_{1 \leq i, j \leq n}$ la matriz de distancias entre estos n individuos.

El objetivo del análisis de conglomerados (o *cluster análisis*) es la **clasificación** (no supervisada) de los elementos de ε , es decir, su agrupación en clases disjuntas, que se denominan **conglomerados** (o *clusters*).

Si estas clases se agrupan sucesivamente en clases de un nivel superior, el resultado es una estructura jerárquica de conglomerados, que puede representarse gráficamente mediante un árbol, llamado **dendogramas**.

Ultramétricas

Se dice que una matriz de distancias D es ultramétrica si para todos los elementos del conjunto ε se verifica que:

$$\left. \begin{array}{l} \delta_{ij} = \delta_{ji}, \text{ para todo } i, j, \\ \delta_{ij} = 0, \text{ para todo } i, \end{array} \right\} \text{disimilitud o casi - métrica}$$

y además verifican la **desigualdad ultramétrica**:

$$\delta_{ij} \leq \max \{ \delta_{ik}, \delta_{kj} \}, \text{ para todo } i, j, k.$$

Correlación Cofenética

Como ocurría en el caso euclídeo, una matriz de distancias D , obtenida a partir de una matriz de datos multivariantes X , no cumple la propiedad ultramétrica.

Esto da lugar al problema de aproximar la matriz de distancias $D = (\delta_{ij})$ con una matriz ultramétrica $U = (u_{ij})$ según algún criterio de proximidad adecuado.

La medida de proximidad que se utiliza es la **correlación cofenética**, que es el coeficiente de correlación lineal (de Pearson) entre los $n(n-1)/2$ pares de distancia (δ_{ij}, μ_{ij}) , para $1 \leq i < j \leq n$.

Este coeficiente vale 1 cuando ambas matrices son proporcionales (iguales). Esto equivale a decir que la matriz D ya cumple la propiedad ultramétrica y, por tanto, la clasificación es exacta.

3.14. VECTOR MEDIO MUESTRAL MULTIVARIADA

Cada elemento es calculado a partir de las n mediciones muestrales de una variable aleatoria. Si X , la matriz de datos, es de dimensión $n \times p$ habrá p medias muestrales.

$$\bar{x}_j = \sum_{i=1}^n \frac{x_{ij}}{n} \text{ para } j = 1, \dots, p$$

Si $x_j = (x_{1j}, x_{2j}, \dots, x_{nj})$ es el j -ésimo vector columna de la matriz de datos conteniendo los valores de las n observaciones de la j -ésima variable, se requieren n ejes para graficar la variable j -ésima en R^n . Suponga, en el mismo espacio el vector con todos los elementos iguales a uno, $\mathbf{1}$. El vector $\mathbf{1}$

n-dimensional normalizado (vector con longitud igual a 1) es $\frac{1}{\sqrt{n}}\mathbf{1}$. Este vector tiene igual ángulo con los n ejes coordenados. La media muestral de una variable representa la proyección del vector de observaciones para esa variable x , sobre el vector unitario normalizado. Luego podemos decir que la media muestral es geoméricamente un múltiplo del vector unitario en el espacio de las observaciones.

$$\begin{aligned} \text{Proy} \left(x \rightarrow \frac{1}{\sqrt{n}}\mathbf{1} \right) &= \frac{x' \cdot (1/\sqrt{n}) \cdot \mathbf{1}}{((1/\sqrt{n}) \cdot \mathbf{1})' \cdot ((1/\sqrt{n}) \cdot \mathbf{1})} (1/\sqrt{n}) \cdot \mathbf{1} \\ &= x' \cdot ((1/\sqrt{n}) \cdot \mathbf{1}) \cdot ((1/\sqrt{n}) \cdot \mathbf{1}) = \frac{1}{n} (x' \cdot \mathbf{1}) \mathbf{1} = \frac{1}{n} (x_1 + x_2 + \dots + x_n) \mathbf{1} = \bar{X} \mathbf{1} \end{aligned}$$

Siempre que se hable de proyecciones de un vector sobre un vector unitario, el resultado de dicha proyección será la media de ese vector.

El vector $\mathbf{x} - \bar{x}\mathbf{1}$ que es ortogonal a la proyección realizada, contiene los residuos o desviaciones de las observaciones con respecto a su media.

El vector $\bar{\mathbf{x}}_p$ que contiene las p medias muestrales puede obtenerse algebraicamente desde la matriz \mathbf{x} de datos, $\bar{\mathbf{x}}_p = \mathbf{x}' \cdot \frac{1}{n} \mathbf{1}$. Este vector media estima el vector esperanza p -dimensional si asumimos que cada observación tiene probabilidad $1/n$ debido al proceso de muestreo aleatorio.

3.15. ANÁLISIS DE COMPONENTES PRINCIPALES

El análisis de componentes principales (ACP) provee una aproximación para la construcción de estas nuevas variables y para decidir cuántas variables podrían ser necesarias para representar información original.

Diferencias entre los datos generan variabilidad. Luego una forma de resumir y ordenar los datos es a través del análisis o la explicación de la estructura de varianza-covarianza del conjunto de variables en estudio. El análisis de componentes principales es una técnica frecuentemente utilizada para ordenar y representar datos multivariados continuos a través de un conjunto de $d = 1, \dots, p$ combinaciones lineales ortogonales normalizadas de las variables originales que explican la variabilidad existente en los datos de forma tal que ningún otro conjunto de combinaciones lineales de igual cardinalidad tiene varianza de las combinaciones mayor a la del conjunto de componentes principales. Usualmente d es mucho menor que p y por ello la técnica implica una reducción de la dimensión y una consecuente facilitación en la interpretación de los datos. No es necesario asumir ninguna estructura a priori del análisis, la idea es encontrar combinaciones lineales de las variables originales con ciertas propiedades de optimalidad (máxima varianza).

El ACP opera sobre la matriz de covarianzas de los datos y preserva la distancia Euclídea, entre observaciones. Los datos podrían o no ser previamente centrado y/o escalados dando lugar a distintos tipos de ACP. El ACP a partir de la matriz de correlación de datos (matriz de covarianzas de las variables originales centradas y escaladas) es útil cuando las unidades de medidas y/o las varianzas de las variables son diferentes, de otro modo las variables de mayor varianza (no necesariamente más informativas) tendrán demasiada influencia en la determinación de la solución. Cuando las variables no tienen varianzas similares o no son medidas sobre la misma escala (variables no conmensurables), la obtención de las componentes a partir de la matriz de correlación es recomendable. Las componentes principales obtenidas usando la matriz de correlación pueden ser sustancialmente diferentes a las obtenidas usando la matriz de covarianzas, en cada caso habrá que juzgar que análisis es más conveniente. Variaciones de la técnica incluyen la asignación de pesos diferenciales para las distintas variables con la finalidad de hacer algunas de ellas más importantes en la descripción de la información (Moser, 1992).

EL ACP no sólo puede realizarse sobre la matriz de covarianza de las p variables de un conjunto de datos multivariados sino también sobre la matriz de covarianza de las n observaciones a los fines de ordenar las variables en el espacio de las componentes resultantes. El ACP para ordenar variables se

basa en la descomposición espectral de la matriz de covarianza o de correlación de dimensión $n \times n$. Las figuras Biplots descritos anteriormente son utilizadas para conectar estos dos análisis.

El ACP para ordenar observaciones se basa en la descomposición espectral de Σ la matriz $p \times p$ de varianzas y covarianzas asociada con el vector p -dimensional de variables aleatorias.

$$\Sigma = \sum_{j=1}^p \lambda_j e_j e_j' = V D_{\lambda} V'$$

Donde v es la matriz de autovectores normalizados y D es la matriz diagonal de autovalores (no negativos) en orden creciente. En la práctica, cuando se trabaja con muestras Σ es reemplazada por s , pero el procedimiento y los resultados son equivalentes, para diferencias la situación componentes derivadas a partir de s suelen llamarse componentes principales muestrales, obviamente estas resumen variabilidad muestral. La selección entre el estimados insesgado y el estimador máximo verosímil de la matriz de covarianza poblacional es irrelevante ya que produce las mismas componentes principales muestrales. Los autovalores y autovectores de la matriz de covarianza muestral podrían ser usados para obtener las componentes muestrales, podrían diferir de su contraparte poblacional debido al muestreo. En adelante no se harán distinción entre autovalores, autovectores, componentes muestrales y poblacionales ya que los resultados

presentados se aplican de igual manera al caso muestral que en el poblacional.

Geoméricamente, las combinaciones lineales representan la selección de un nuevo sistema de coordenadas obtenidos por la rotación del sistema de coordenadas original. Los nuevos ejes representan las direcciones con máxima variabilidad y proveen una forma más simple y parsimoniosa para describir la estructura de covarianza. Esta nueva base para representar los datos satisface que el primer eje se encuentra en dirección de máxima varianza (la primera componente es la combinación lineal de máxima varianza) y los ejes subsecuentes maximizan la varianza explicada dado que son ortogonales a los ejes previos.

Dado que las componentes son ortogonales, son no correlacionadas. Luego, la proporción de la varianza total, traza de la matriz de covarianzas, explicada por las primeras d nuevas variables será:

$$\text{Pr op}_d = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_d}{\lambda_1 + \lambda_2 + \dots + \lambda_p}$$

Con Pr op_d tan grande como sea posible para los datos analizados. (La varianza total en la muestra será igual a p cuando se trabaje con variables previamente estandarizadas).

Si bien el número de componentes principales posibles de construir es p , para obtener una dimensión de reducción simplemente se selecciona un orden d menor a p de combinaciones lineales que retiene una adecuada proporción de la varianza total (digamos el 75% o más) y se usan estas combinaciones como nuevas variables para graficar y analizar los datos sin mayor pérdida de información.

Bajo normalidad multivariada, los componentes principales muestrales debieran distribuirse normalmente, el análisis de figuras de dispersión univariados (Q-Q plots para cada componente principal) y bivariados (diagramas de dispersión de los primeros pares de componentes) pueden ser utilizados como otro mecanismo para la verificación de la normalidad multivariada en los datos.

La construcción de éstos figuras implica la obtención de

$D_i^2 = (x_i - \bar{x})' S^{-1} (x_i - \bar{x})$. Las componentes principales pueden ser utilizadas para obtener muy fácilmente este estadístico; si cada vector de observación p dimensional x_i , es transformado reemplazando el elemento j -ésimo por la j -ésima componente principal estandarizada, y llamando z_i al vector

$$\text{transformado, entonces } D_i^2 = z_i' z_i = \sum_{j=1}^p z_{ij}^2 .$$

Las figuras de dispersión de las últimas componentes principales muestrales pueden utilizarse para detectar observaciones sospechosas o aberrantes. Cada vector de observación puede reconstruirse desde las componentes principales y el conjunto completo de autovectores S , como:

$$x_i = (x'_1 e_1) e_1 + (x'_2 e_2) e_2 + \dots + (x'_p e_p) e_p$$

$$x_i = y_{1i} e_1 + y_{2i} e_2 + \dots + y_{pi} e_p$$

Luego, las magnitudes de las últimas componentes principales determinan cuan bien las primeras ajustan a las observaciones.

Las componentes principales pueden ser utilizadas también para estudiar diferencias entre grupos cuando los datos se encuentran clasificados ya que las diferencias entre grupos típicamente induce variabilidad en las mediciones.

Para analizar asociación entre variables con componentes. Las correlaciones entre las componentes principales y las variables originales viene dada por:

$$r(Y_j, X_k) = \frac{e_{kj} \sqrt{\lambda_j}}{\sqrt{\sigma_k^2}}$$

Y representan un indicados de cuan importante es una variable particular en la construcción de la componente. La interpretación de esta correlación puede ser más confiable que la interpretación de los coeficientes de la combinación ya que la correlación tiene en cuenta diferencias en las

varianzas de las variables originales y consecuentemente elimina el sesgo de interpretaciones causadas por diferentes escalas de medición.

3.16. ANÁLISIS DE CORRESPONDENCIAS

El análisis de correspondencia (AC) es una técnica exploratoria que permite representar gráficamente filas y columnas de una tabla de contingencia (Greenacre, 1984, 1988, 1994; Lebart et al, 1984). En Psicología suelen referirse a esta técnica como escalamiento dual; en Ecología ha sido muy usada para ordenamiento de datos discretos de vegetación (presencia-ausencia de una serie de especies en cada parcela observada a lo largo de un gradiente ambiental). La técnica de AC también constituye una herramienta de principal importancia para el análisis de datos textuales donde se construyen tablas de contingencia relacionando el uso de varios vocablos entre distintos textos de discurso. El AC puede ser interpretado como una técnica complementaria y a veces suplementaria del uso de modelos log-lineales para el estudio analítico de las relaciones contenidas en tablas de contingencia. El AC permite explorar gráficamente estas relaciones.

En el AC se representan filas y las columnas de una tabla a dos vías de variables categorizadas, como puntos en un espacio Euclídeo de baja dimensión (generalmente bidimensional). El propósito de su uso es similar al

del análisis de componentes principales para datos continuos, diferenciándose de este por el hecho de que el AC opera sobre la matriz de desviaciones Chi cuadrado en lugar de usar una matriz de varianzas y covarianzas.

Las filas de las tablas de contingencia pueden ser vistas como puntos con coordenadas dadas por las columnas de la tabla. Los perfiles filas son construidos a partir de la división de la frecuencia observada en cada celda por el correspondiente total de fila. A cada punto fila se le asigna un peso a través de la división del total de fila por el gran total de la tabla. Los perfiles columnas se definen de manera equivalente. El AC determina, a través de la descomposición por valor singular de la matriz de desviaciones ji cuadrado de proporciones filas y columnas bajo la hipótesis de independencia entre filas y columnas, un subespacio óptimo para la representación de los perfiles filas y columnas ponderadas por sus respectivos pesos.

Cuando el AC es realizado sobre una única tabla de dos vías, se denomina *Análisis de Correspondencia Simple (ACS)*. Este análisis permite graficar observaciones bivariadas en planos e identificar las asociaciones de mayor peso entre las modalidades de dos variables cualitativas. El *Análisis de Correspondencia Múltiple (ACS)* permite explorar tablas multidimensionales. Las observaciones multivariadas se grafican en planos para así poder

identificar las asociaciones de mayor peso entre las modalidades de varias variables cualitativas. Para este último enfoque se utilizan las conocidas tablas de Burt que contienen los niveles o modalidades de cada variable categorizada tanto en las filas como en las columnas de la tabla y por tanto contienen todas las clasificaciones cruzadas a dos vías de las variables originales.

El AC opera sobre la matriz de desviaciones χ^2 cuadrado (mencionada en el Capítulo III), en lugar de usar la matriz de varianzas y covarianzas como lo hace el análisis de componentes principales. Este método mide cuáles son las combinaciones de modalidades que tienen más inercia (que más contribuyen a rechazar la hipótesis de independencia entre las dos variables). Ellas son las modalidades de la periferia o modalidades que se alejan del centro del plano. Como el análisis no es realizado sobre las frecuencias absolutas sino sobre las proporciones de la tabla de contingencia, comúnmente se utiliza el término inercia para denotar la información χ^2 cuadrado en la tabla (inercia es el valor χ^2 cuadrado dividido por el gran total de la tabla).

Para la matriz de desviaciones por celda se obtiene un conjunto de autovectores y autovalores que son usados para construir un subespacio óptimo para la representación de los perfiles filas y columnas ponderadas por

sus respectivos pesos. Los ejes son extraídos en relación a la desviación χ^2 cuadrado explicada por cada uno. El primer eje principal se asocia a la más alta contribución sobre el estadístico χ^2 cuadrado de la tabla de contingencia. Los primeros d ejes definen el espacio d -dimensional óptimo, con $d = \min(i - 1, j - 1)$ con $i = \text{número de filas}$, $j = \text{números de columnas}$. La proporción de la inercia total explicada por cada eje es usada como criterio de selección del número de ejes necesarios para la representación.

Similar al análisis de componentes principales, los resultados pueden ser representados en un Biplot para graficar los puntos filas y columnas en el mismo espacio. Las distancias entre éstos puntos filas miden la discrepancia entre perfiles filas. Los puntos filas muy cercanos en el figura, tienen similar perfil fila. Distancias desde el origen indican la discrepancia entre los perfiles filas y el centroide fila o la distribución fila marginal. El mismo tipo de interpretación puede realizarse sobre los perfiles columnas. Las distancias entre puntos filas y columnas carecen de sentido, pero puntos filas y columnas que caen en la misma dirección respecto al origen se encuentran positivamente correlacionados, mientras que aquellos que caen en direcciones opuestas se encuentran negativamente correlacionados. Las direcciones pueden cambiar si se grafican otras dimensiones, por lo que es importante realizar el análisis sobre un espacio con alta inercia.

CAPÍTULO IV

ANÁLISIS ESTADÍSTICO

INTRODUCCIÓN

En éste capítulo presentaremos el análisis estadístico univariado y multivariado de datos recolectados de la planta de banano en la hacienda Carmita en Balao para el estudio de dos grupos de variables mencionados en el capítulo II, las relacionadas al estado evolutivo de la Sigatoka Negra y las relacionadas con el clima. La hacienda se encuentra clasificada en cuatro lotes, Lote A, Lote B, Lote C y Lote E.

Los datos de las variables para el análisis han sido registrados semanalmente en la bananera, analizaremos desde la primera semana del mes de abril del 2005 hasta la última semana del mes de diciembre del 2006.

Los datos de las variables relacionadas con el clima fueron capturados cada media hora durante las veinticuatro horas desde el primer día del mes de abril del 2005 hasta el 31 de diciembre del 2006.

4.2. ANÁLISIS DE VARIABLES

Para el análisis de las variables se utilizaron los siguientes programas estadísticos: MINITAB 13.0 para Windows, SPSS 12.0 e InfoStat 2.0.

Las variables que analizaremos de *Preaviso* de la enfermedad son: Hoja III (nivel de infección en hoja 3), Hoja IV nivel de infección en hoja 4), EE (Estado Evolutivo); de *Severidad* de la enfermedad son: Floración (nivel de infección de la planta a la floración) y Cosecha (nivel de infección de la planta a la cosecha); y las variables relacionados con el clima tomaremos las determinadas por Ing. Chóez Salazar M.^[7] y éstas son: Rain, Temp Out, Temp Hi, Temp Low, Wind Chill, Heat Index, THW Index, THSW Index, Cool D-D, Solar Rad, Solar Energy, Hi Solar Energy, UV Index, UV Dose, Hi UV, ET, Out Hum, Dew Pt., Wind Speed.

Con los datos obtenidos diariamente se procedió a realizar el cálculo de la Acumulada y el Promedio mensual, ejemplo: Rain Acumulada y Rain Promedio, que sería Lluvia Acumulada (Precipitación) y Lluvia Promedio, Temp Out Acumulada y Temp Out Promedio, es la Temperatura Exterior Acumulada y Temperatura Exterior Promedio y de igual manera con las demás variables.

4.3. ANÁLISIS UNIVARIADO

A continuación realizaremos el estudio univariado de las variables relaciones con la enfermedad y las relacionadas con el clima analizando éstas últimas con su acumulada y su promedio.

4.3.1. VARIABLES DE PREAVISO

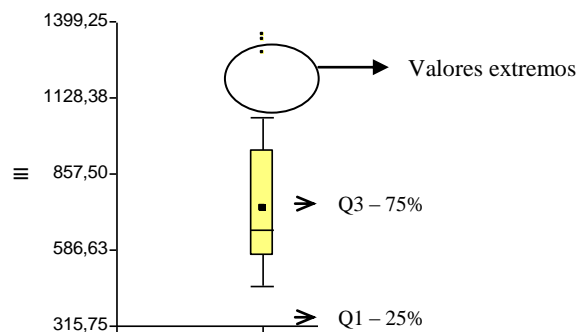
4.3.1.1. HOJA III

La variable III nos indica el estado de la enfermedad en la etapa de preaviso en la Hoja 3 de la planta de banano.

Se obtuvo una media de 732,67 con una desviación estándar de 249,43 de evolución de la enfermedad un valor mínimo de 365 y un valor máximo de 1350. El tamaño de la muestra es de 52 datos.

En la figura 4.1 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 560,5 y de igual manera en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 942,5. Además podemos observar que en ésta variable existen valores extremos.

Figura 4.1 Diagrama de Cajas del Estado de la Enfermedad en la Hoja III



Con los datos obtenidos de asimetría 0.79 decimos que la curva es *asimétricamente positiva* por lo que los valores se tienden a reunir más en la parte izquierda de la media (véase figura 4.2), en cuanto a la puntiagudez respecto a la normal la distribución es *mesocúrtica* ya que el coeficiente de curtosis es -0.12 encontrándose entre ± 0.5 .

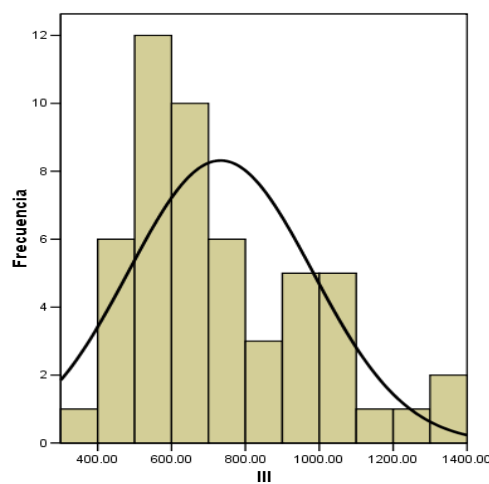


Figura 4.2 Histograma del Estado de la Enfermedad en la Hoja III

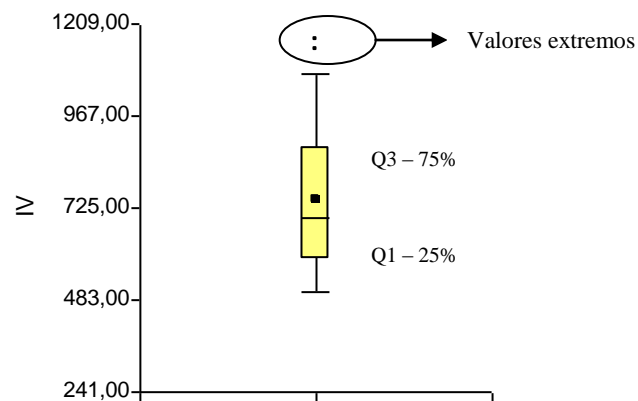
4.3.1.2. HOJA IV

La variable IV nos indica el estado de la enfermedad en la Hoja 4 de la planta de banano.

Se obtuvo una media de 744,02 con una desviación estándar de 213,81 de evolución de la enfermedad un valor mínimo de 285 y un valor máximo de 1165. El tamaño de la muestra es de 52 datos.

En la figura 4.3 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 589 y de igual manera en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 882,5. Además podemos observar que en ésta variable existen valores extremos.

Figura 4.3 Diagrama de Cajas del Estado de la Enfermedad en la Hoja IV



Con los datos obtenidos de asimetría 0.38 decimos que la curva es *simétrica* ya que se encuentra entre (± 0.5) es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media (véase figura 4.4), en cuanto a la puntiagudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es negativo -0.61.

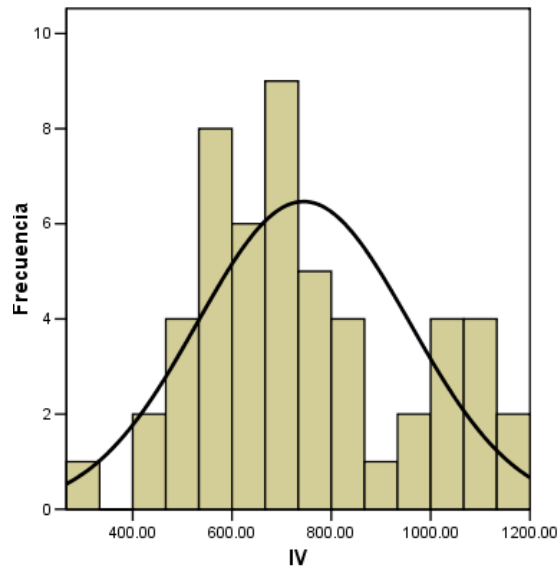


Figura 4.4 Histograma del Estado de la Enfermedad en la Hoja IV

4.3.1.3. ESTADO EVOLUTIVO

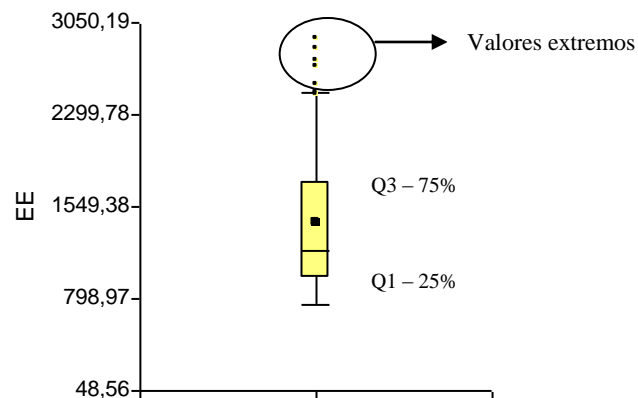
La variable EE nos indica el estado evolutivo de la enfermedad en la planta de banano.

Se obtuvo una media de 1414,07 con una desviación estándar de 656,88 de evolución de la enfermedad un valor mínimo de 185 y un valor máximo de 2913,75. El tamaño de la muestra es de 52 datos.

En la figura 4.5 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 955,23 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 1748,5. Además podemos observar que en ésta variable existen valores extremos.

Figura 4.5 Diagrama de Cajas del Estado Evolutivo



Con los datos obtenidos de asimetría 0.71 decimos que la curva es *asimétricamente positiva* por lo que los valores se tienden a reunir más en la parte izquierda de la media (véase figura 4.6), en cuanto a la puntiagudez respecto a la normal la distribución es *mesocúrtica* ya que el coeficiente de curtosis es negativo -0.2 y encontrándose entre (± 0.5) .

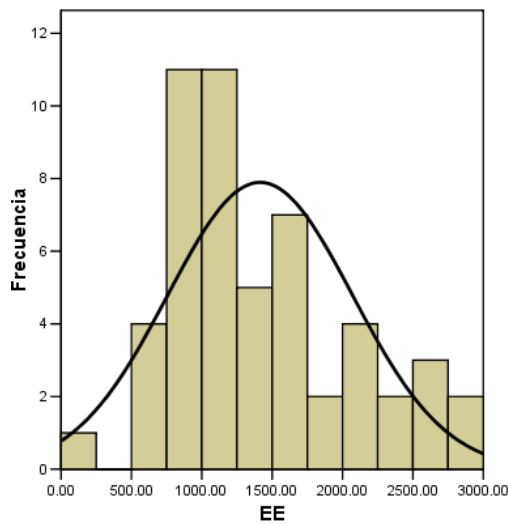


Figura 4.6 Histograma del Estado Evolutivo

4.3.2. SEVERIDAD

4.3.2.1. FLORACIÓN

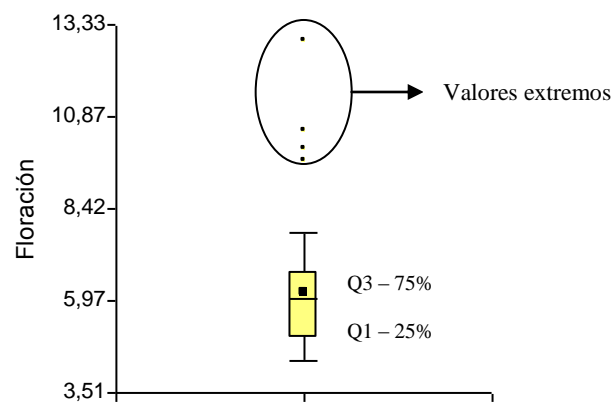
Esta variable nos indica la severidad de la enfermedad de la planta en su etapa de floración.

Se obtuvo una media de 6,17 con una desviación estándar de 1,7 de severidad de la enfermedad un valor mínimo de 3,96 y un valor máximo de 12,88. El tamaño de la muestra es de 52 datos.

En la figura 4.7 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 4,98 y de igual manera en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o

iguales a 6,71. Además podemos observar que en ésta variable existen valores extremos.

Figura 4.7 Diagrama de Cajas de la Enfermedad en la Floración



Con los datos obtenidos de asimetría 1.74 decimos que la curva es *asimétricamente positiva* por lo que los valores se tienden a reunir más en la parte izquierda de la media (véase figura 4.8), en cuanto a la puntiagudez respecto a la normal la distribución es *leptocúrtica* ya que el coeficiente de curtosis es positivo 4,52.

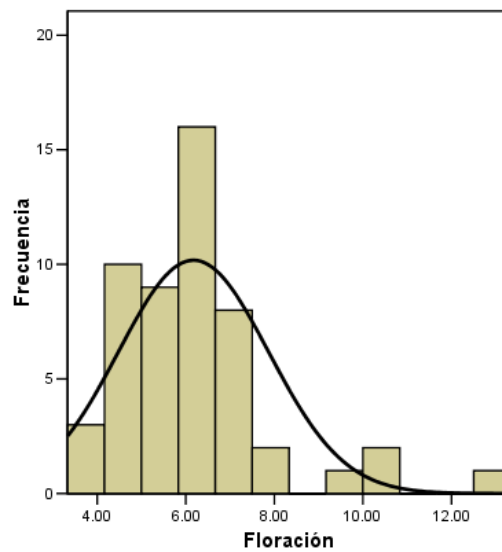


Figura 4.8 Histograma del Estado de la Enfermedad en Floración

4.3.2.2. COSECHA

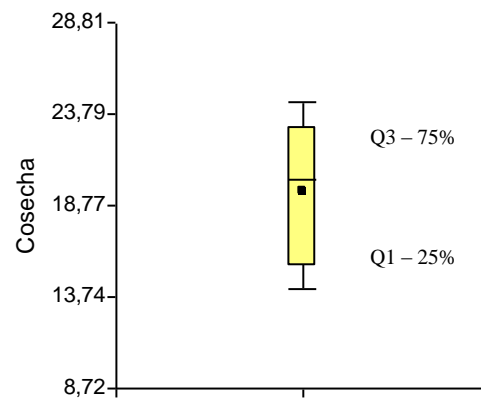
La variable cosecha nos indica la severidad de la enfermedad de la planta en la etapa de la recolección de la fruta o cosecha.

Se obtuvo una media de 19,5 con una desviación estándar de 4,39 de severidad de la enfermedad un valor mínimo de 9,63 y un valor máximo de 27,9. El tamaño de la muestra es de 52 datos.

En la figura 4.9 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 15,43 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 23,03. En ésta variable no existen valores extremos.

Figura 4.9 Diagrama de Cajas del Estado de la Enfermedad en Cosecha



Con los datos obtenidos de asimetría $-0,17$ decimos que la curva es *simétrica* ya que se encuentra entre (± 0.5) es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media (véase figura 4.10), en cuanto a la puntigudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es negativo $-0,78$.

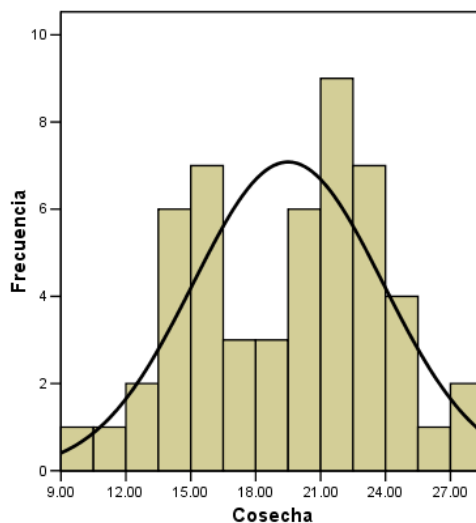


Figura 4.10 Histograma del Estado de la Enfermedad en la Cosecha

4.3.3. CLIMA

4.3.3.1. RAIN ACUMULADA

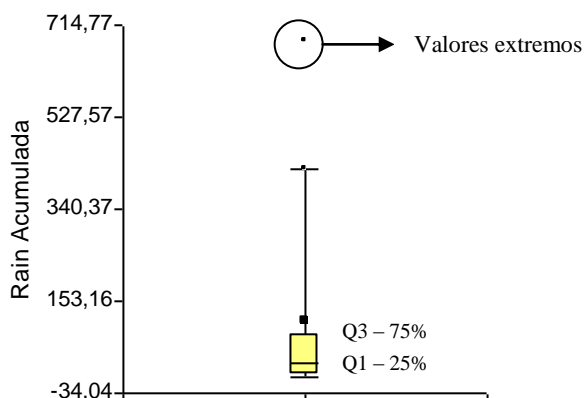
La variable Rain Acumulada nos indica la cantidad de lluvia acumulada por mes registrada en mm también la podemos llamar Precipitación.

Se obtuvo una media de 112,18 mm con una desviación de 199,18 mm de precipitación un valor mínimo de 0 y un valor máximo de 680,73, el valor mínimo que toma la variable se debe que en algunos meses no llueve porque es la estación de verano. El tamaño de la muestra es de 52 datos.

En la figura 4.11 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 2,03 y de igual manera en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o

iguales a 85,28. Además podemos observar que en ésta variable existen valores extremos.

Figura 4.11 Diagrama de Cajas de Rain Acumulada



Con los datos obtenidos de asimetría 2,11 decimos que la curva es *asimétricamente positiva* por lo que los valores se tienden a reunir más en la parte izquierda de la media (véase figura 4.12), en cuanto a la puntiagudez respecto a la normal la distribución es *leptocúrtica* ya que el coeficiente de curtosis es positiva con un valor de 3,24.

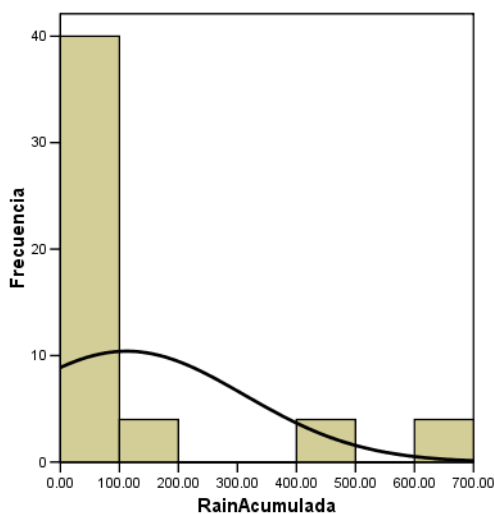


Figura 4.12 Histograma de Rain Acumulada

4.3.3.2. RAIN PROMEDIO

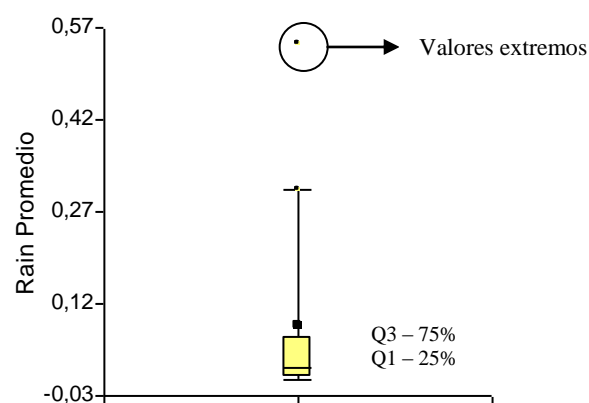
La variable Rain Promedio nos indica la cantidad de lluvia promedio por mes registrada en mm.

Se obtuvo una media de 0,09 mm con una desviación de 0,16 de lluvia mensual un valor mínimo de 0 y un valor máximo de 0,55, el valor mínimo que toma la variable se debe que en algunos meses no llueve porque es la estación de verano. El tamaño de la muestra es de 52 datos.

En la figura 4.13 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 0 y de igual manera en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a

0,07. Además podemos observar que en ésta variable existen valores extremos.

Figura 4.13 Diagrama de Cajas de Rain Promedio



Con los datos obtenidos de asimetría 2,22 decimos que la curva es *asimétricamente positiva* por lo que los valores se tienden a reunir más en la parte izquierda de la media (véase figura 4.14), en cuanto a la puntiagudez respecto a la normal la distribución es *leptocúrtica* ya que el coeficiente de curtosis es positiva con un valor de 3,79.

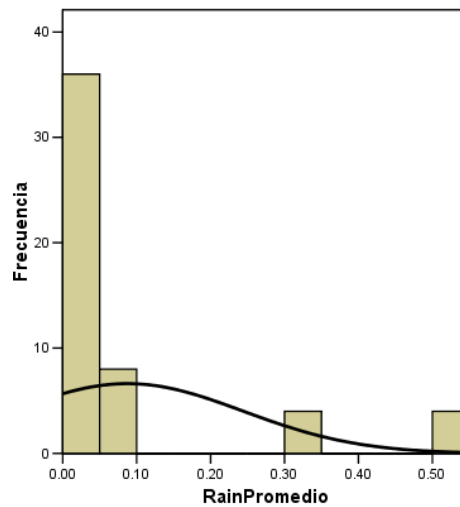


Figura 4.14 Histograma de Rain Promedio

4.3.3.3. TEMP OUT ACUMULADA

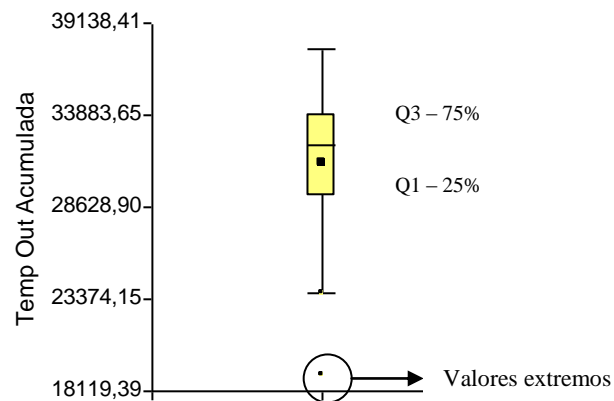
La variable Temp Out Acumulada nos representa la Temperatura Exterior Acumulada expresada en grados centígrados (°C).

Se obtuvo una media de 31180,43 con una desviación estándar de 5125,45 un valor mínimo de 19074,8 y un valor máximo de 38183. El tamaño de la muestra es de 52 datos.

En la figura 4.15 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 29239,8 y de igual manera en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o

iguales a 33833,4. Además podemos observar que en ésta variable existen valores extremos.

Figura 4.15 Diagrama de Cajas de Temp Out Acumulada



Con los datos obtenidos de asimetría $-0,88$ decimos que la curva es *asimétricamente negativa* por lo que los valores se tienden a reunir más en la parte derecha de la media (véase figura 4.16), en cuanto a la puntiagudez respecto a la normal la distribución es *mesocúrtica* ya que el coeficiente de curtosis es positiva con un valor de $0,47$ y encontrándose entre (± 0.5) .

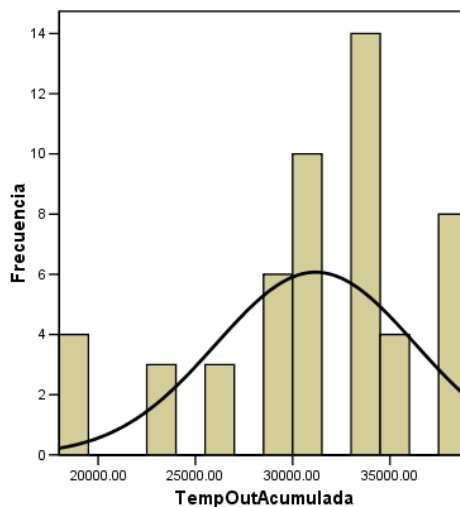


Figura 4.16 Histograma de Temp Out Acumulada

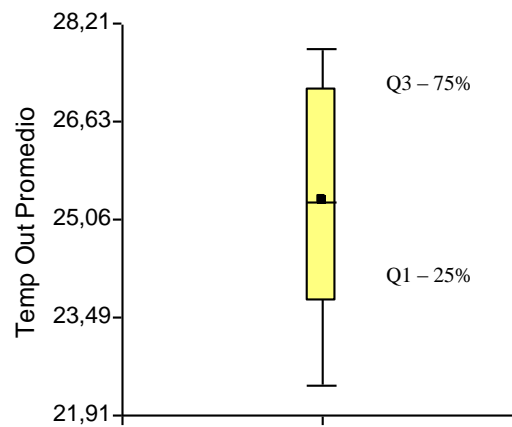
4.3.3.4. TEMP OUT PROMEDIO

La variable Temp Out Promedio nos representa la Temperatura Exterior Promediada y expresada ésta en grados centígrados (°C).

Se obtuvo una media de 25,36 con una desviación estándar de 1,92 un valor mínimo de 22,2 y un valor máximo de 27,92. El tamaño de la muestra es de 52 datos.

En la figura 4.17 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 23,71 y de igual manera en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 27,13. En ésta variable no existen valores extremos.

Figura 4.17 Diagrama de Cajas de Temp Out Promedio



Con los datos obtenidos de asimetría $-0,19$ decimos que la curva es *simétrica* ya que se encuentra entre (± 0.5) , es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media (véase figura 4.18), en cuanto a la puntigudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es negativo con un valor de $-1,31$.

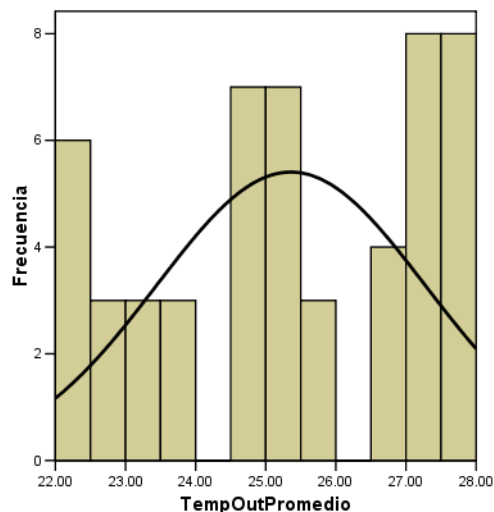


Figura 4.18 Histograma de Temp Out Promedio

4.3.3.5. TEMP HI ACUMULADA

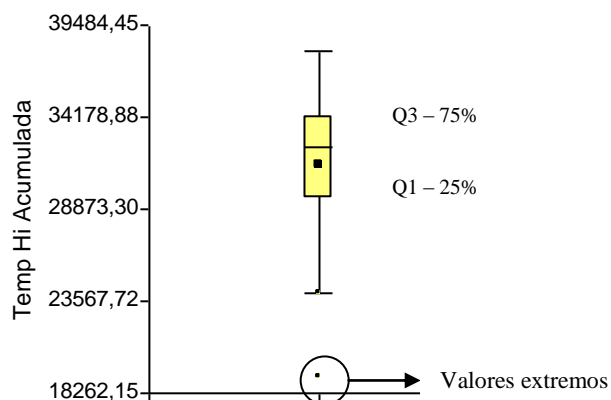
La variable Temp Hi Acumulada nos representa la Temperatura Máxima Exterior Acumulada y expresada en grados centígrados (°C).

Se obtuvo una media de 31444,21 con una desviación estándar de 5159,42 un valor mínimo de 19226,8 y un valor máximo de 38519,8. El tamaño de la muestra es de 52 datos.

En la figura 4.19 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 29484 y de igual manera en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o

iguales a 34122,1. Además podemos observar que en ésta variable existen valores extremos.

Figura 4.19 Diagrama de Cajas de Temp Hi Acumulada



Con los datos obtenidos de asimetría $-0,89$ decimos que la curva es *asimétricamente negativa* por lo que los valores se tienden a reunir más en la parte derecha de la media (véase figura 4.20), en cuanto a la puntiagudez respecto a la normal la distribución es *mesocúrtica* ya que el coeficiente de curtosis es positivo con un valor de $0,49$ y encontrándose entre (± 0.5) .

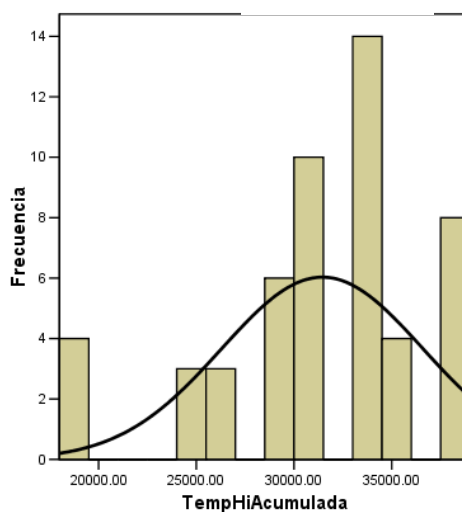


Figura 4.20 Histograma de Temp Hi Acumulada

4.3.3.6. TEMP HI PROMEDIO

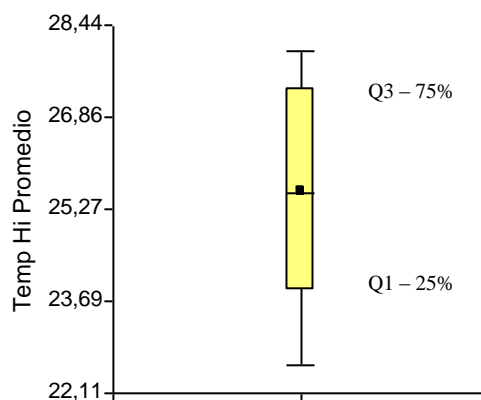
La variable Temp Hi Promedio nos representa la Temperatura Máxima Exterior Promediada y expresada ésta en grados centígrados (°C).

Se obtuvo una media de 25,57 con una desviación estándar 1,92, un valor mínimo de 22,4 y un valor máximo de 28,15. El tamaño de la muestra es de 52 datos.

En la figura 4.21 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 22,4 y de igual manera en

el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 28,15. En ésta variable no existen valores extremos.

Figura 4.21 Diagrama de Cajas de Temp Hi Promedio



Con los datos obtenidos de asimetría $-0,19$ decimos que la curva es *simétrica* ya que se encuentra entre (± 0.5) , es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media (véase figura 4.22), en cuanto a la puntigudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es negativo con un valor de $-1,31$.

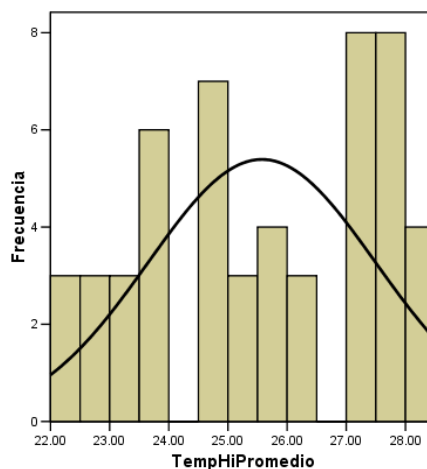


Figura 4.22 Histograma de Temp Hi Promedio

4.3.3.7. TEMP LOW ACUMULADA

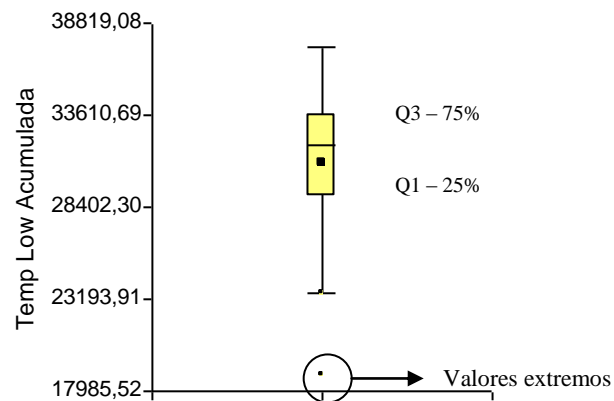
La variable Temp Low Acumulada nos representa la Temperatura Mínima Exterior Acumulada y expresada ésta en grados centígrados (°C).

Se obtuvo una media de 30932,35 con una desviación estándar de 5092,26 un valor mínimo de 18932,5 y un valor máximo de 37872,1. El tamaño de la muestra es de 52 datos.

En la figura 4.23 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 29004,65 y de igual manera en el cuartil 3 (Q_3) el 75% de las observaciones toman valores

menores o iguales a 33556,3. Además podemos observar que en ésta variable existen valores extremos.

Figura 4.23 Diagrama de Cajas de Temp Low Acumulada



Con los datos obtenidos de asimetría $-0,88$ decimos que la curva es *asimétricamente negativa* por lo que los valores se tienden a reunir más en la parte derecha de la media (véase figura 24), en cuanto a la puntiagudez respecto a la normal la distribución es *mesocúrtica* ya que el coeficiente de curtosis es positivo con un valor de $0,45$ y encontrándose entre (± 0.5) .

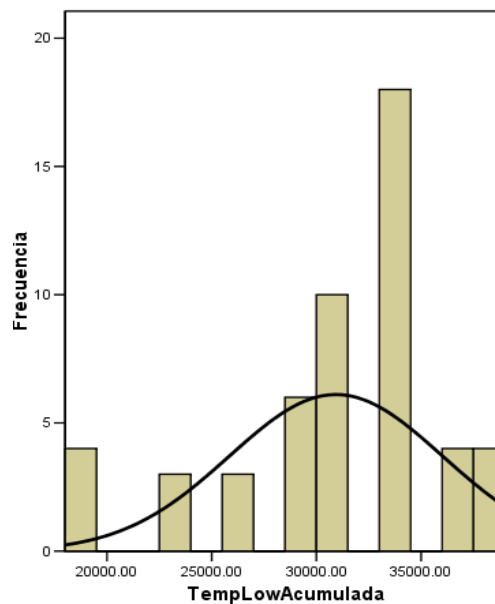


Figura 4.24 Histograma de Temp Low Acumulada

4.3.3.8. TEMP LOW PROMEDIO

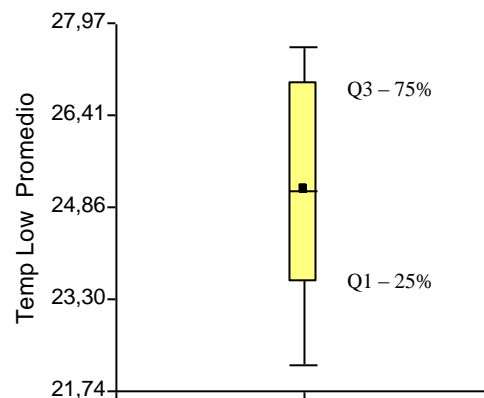
La variable Temp Low Promedio nos representa la Temperatura Mínima Exterior Promediada y expresada ésta en grados centígrados (°C).

Se obtuvo una media de 25,16 con una desviación estándar de 1,91 un valor mínimo de 22,02 y un valor máximo de 27,69. El tamaño de la muestra es de 52 datos.

En la figura 4.25 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 23,55 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 26,94. En ésta variable no existen valores extremos.

Figura 4.25 Diagrama de Cajas de Temp Low Promedio



Con los datos obtenidos de asimetría $-0,19$ decimos que la curva es *simétrica* ya que se encuentra entre (± 0.5) , es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media (véase figura 4.26), en cuanto a la puntiagudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es negativo con un valor de $-1,31$.

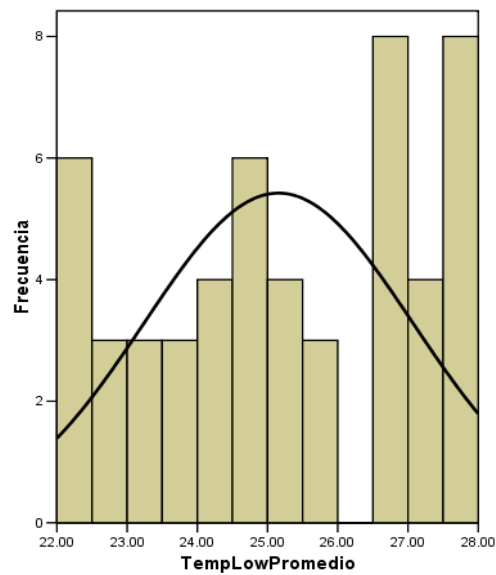


Figura 4.26 Histograma de Temp Low Promedio

4.3.3.9. WIND CHILL ACUMULADA

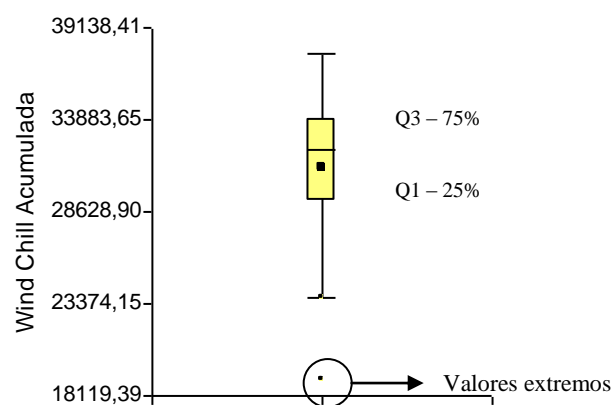
La variable Wind Chill Acumulada nos representa la Temperatura Aérea en relación a la Temperatura del cuerpo humano acumulado y expresado ésta en grados centígrados (°C).

Se obtuvo una media de 31180,43 con una desviación estándar de 5126,45 un valor mínimo de 19074,8 y un valor máximo de 38183. El tamaño de la muestra es de 52 datos.

En la figura 4.27 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 29239,8 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 33833,4. Además podemos observar que en ésta variable existen valores extremos.

Figura 4.27 Diagrama de Cajas de Wind Chill Acumulada



Con los datos obtenidos de asimetría $-0,88$ decimos que la curva es *asimétricamente negativa* por lo que los valores se tienden a reunir más en la parte derecha de la media (véase figura 4.28), en cuanto a la puntiagudez respecto a la normal la distribución es *mesocúrtica* ya que el coeficiente de curtosis es positivo con un valor de $0,47$ y encontrándose entre (± 0.5) .

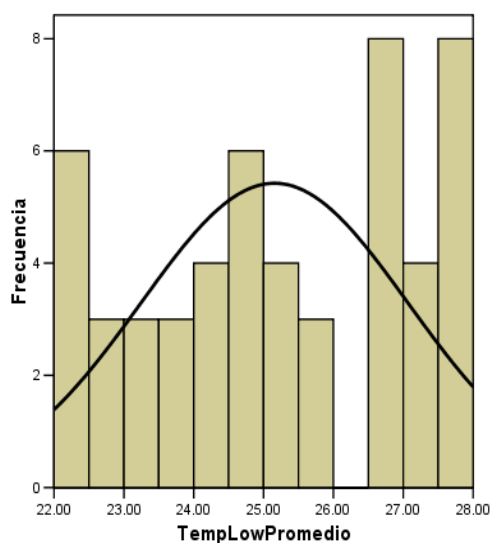


Figura 4.28 Histograma Wind Chill Promedio

4.3.3.10. WIND CHILL PROMEDIO

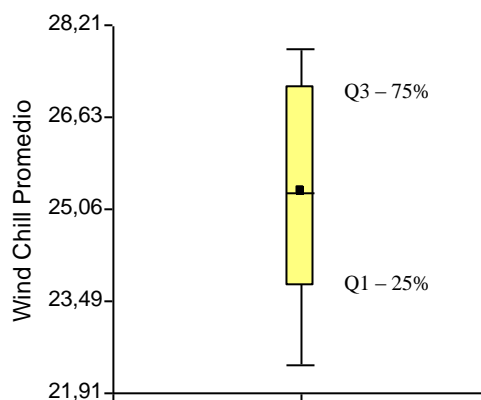
La variable Wind Chill Promedio nos representa la Temperatura Aérea en relación a la Temperatura del cuerpo humano promediado y expresada ésta en grados centígrados (°C).

Se obtuvo una media de 25,36 con una desviación estándar de 1,92 un valor mínimo de 22,2 y un valor máximo de 27,92. El tamaño de la muestra es de 52 datos.

En la figura 4.29 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 23,71 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 27,13. En ésta variable no existen valores extremos.

Figura 4.29 Diagrama de Cajas de Wind Chill Promedio



Con los datos obtenidos de asimetría $-0,19$ decimos que la curva es *simétrica* ya que se encuentra entre (± 0.5) , es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media (véase figura 4.30), en cuanto a la puntiagudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es negativo con un valor de $-1,31$.

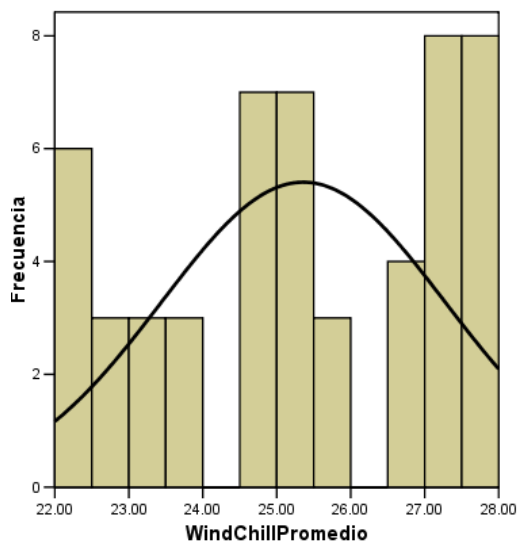


Figura 4.30 Histograma de Wind Chill Promedio

4.3.3.11. HEAT INDEX ACUMULADA

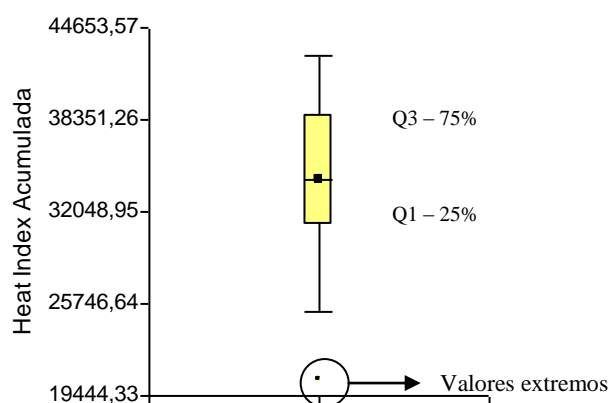
La variable Heat Index Acumulada nos representa el índice de calor o temperatura relacionada a la humedad acumulado y expresado ésta en grados centígrados (°C).

Se obtuvo una media de 34243,44 con una desviación estándar de 6345,55 un valor mínimo de 20590,2 y un valor máximo de 43507,7. El tamaño de la muestra es de 52 datos.

En la figura 4.31 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 20590,2 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales 43507,7. Además podemos observar que en ésta variable existen valores extremos.

Figura 4.31 Diagrama de Cajas Heat Index Acumulada



Con los datos obtenidos de asimetría $-0,53$ decimos que la curva es *asimétricamente negativa* por lo que los valores se tienden a reunir más en la parte derecha de la media (véase figura 4.32), en cuanto a la puntiagudez respecto a la normal la distribución es *mesocúrtica* ya que el coeficiente de curtosis es negativo con un valor de $-0,22$ y encontrándose entre (± 0.5) .

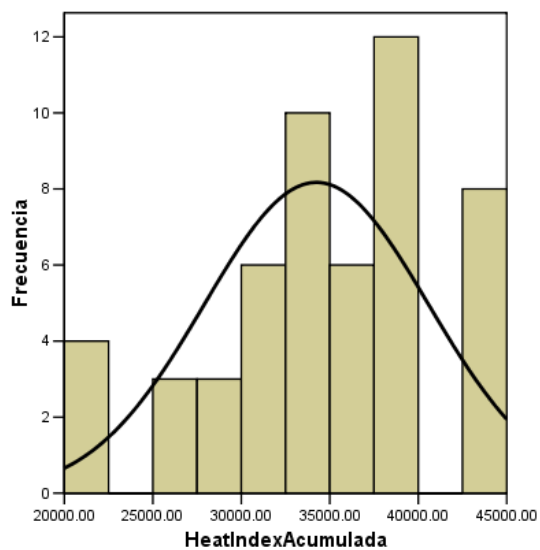


Figura 4.32 Histograma Heat Index Acumulada

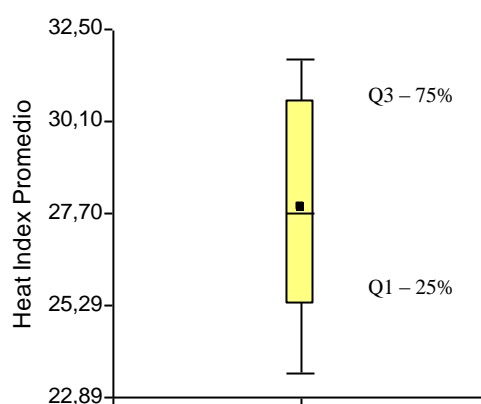
4.3.3.12. HEAT INDEX PROMEDIO

La variable Heat Index Promedio nos representa el índice de calor o temperatura relacionada a la humedad promediado y expresado ésta en grados centígrados (°C).

Se obtuvo una media de 27,83 con una desviación estándar de 2,96 un valor mínimo de 23,33 y un valor máximo de 32,06. El tamaño de la muestra es de 52 datos.

En la figura 4.33 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 25,29 y de igual manera en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 30,62. En ésta variable no existen valores extremos.

Figura 4.33 Diagrama de Cajas de Heat Index Promedio



Con los datos obtenidos de asimetría $-0,01$ decimos que la curva es *simétrica* ya que se encuentra entre (± 0.5) , es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media (véase figura 4.34), en cuanto a la puntigudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es negativo con un valor de $-1,4$.

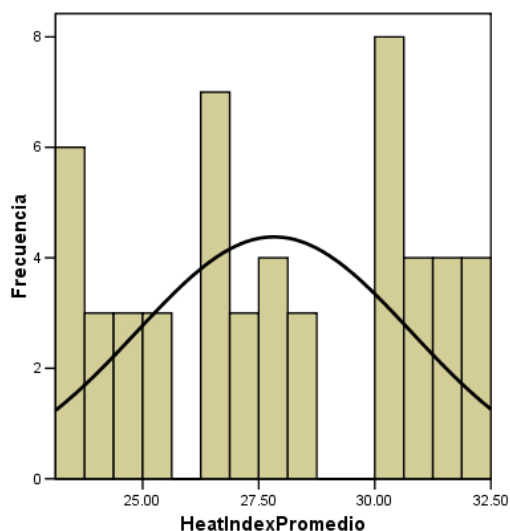


Figura 34 Histograma de Heat Index Promedio

4.3.3.13. THW INDEX ACUMULADA

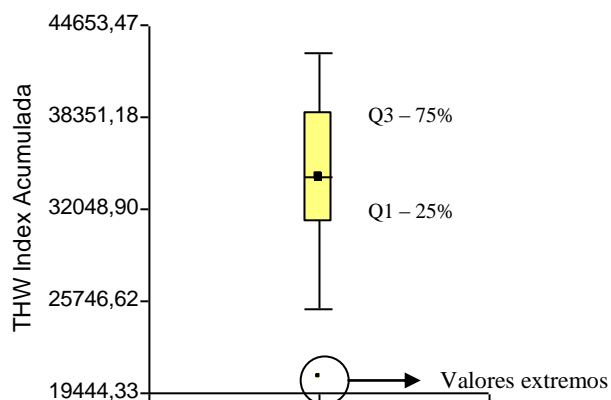
La variable THW Index Acumulada es el índice de calor o temperatura relacionada a la humedad y expresado ésta en grados centígrados (°C).

Se obtuvo una media de 34243,43 con una desviación estándar de 6345,53 un valor mínimo de 20590,2 y un valor máximo de 43507,6. El tamaño de la muestra es de 52 datos.

En la figura 4.35 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 31041,75 y de igual manera en el cuartil 3 (Q_3) el 75% de las observaciones toman valores

menores o iguales a 38672,4. Además podemos observar que en ésta variable existen valores extremos.

Figura 4.35 Diagrama de Cajas de THW Index Acumulada



Con los datos obtenidos de asimetría $-0,53$ decimos que la curva es *asimétricamente negativa* por lo que los valores se tienden a reunir más en la parte derecha de la media (véase figura 4.36), en cuanto a la puntiagudez respecto a la normal la distribución es *mesocúrtica* ya que el coeficiente de curtosis es negativo con un valor de $-0,22$ y encontrándose entre (± 0.5) .

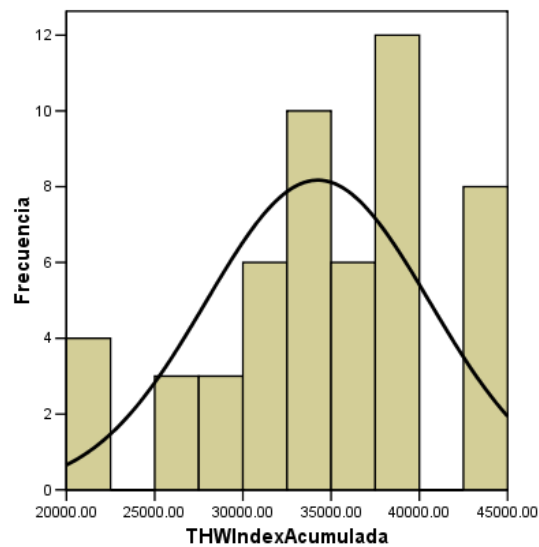


Figura 4.36 Histograma de THW Index Acumulada

4.3.3.14. THW INDEX PROMEDIO

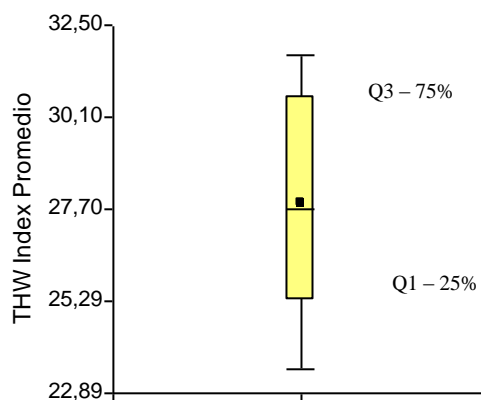
La variable THW Index Promedio es el índice de calor o temperatura relacionada a la humedad promediado y expresado ésta en grados centígrados (°C).

Se obtuvo una media de 27,83 con una desviación estándar de 2,96 un valor mínimo de 23,33 y un valor máximo de 32,06. El tamaño de la muestra es de 52 datos.

En la figura 4.37 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 23,33 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 32,06. En ésta variable no existen valores extremos.

Figura 4.37 Diagrama de Cajas de THW Index Promedio



Con los datos obtenidos de asimetría $-0,01$ decimos que la curva es *simétrica* ya que se encuentra entre (± 0.5) , es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media (véase figura 4.38), en cuanto a la puntiagudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es negativo con un valor de $-1,4$

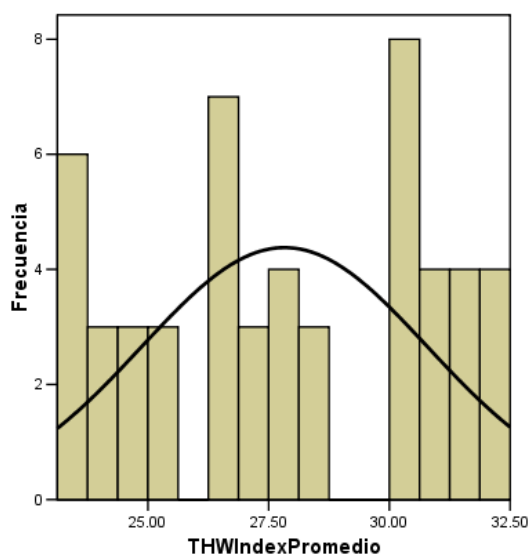


Figura 4.38 Histograma de THW Index Promedio

4.3.3.15. THSW INDEX ACUMULADA

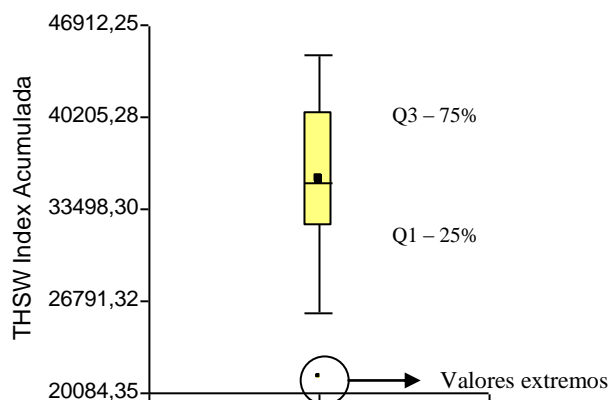
La variable THSW Index Acumulada es el índice que relaciona la temperatura, humedad, radiación solar y viento acumulado y expresado ésta en grados centígrados (°C).

Se obtuvo una media de 35644,79 con una desviación estándar de 6781,92 un valor mínimo de 21303,8 y un valor máximo de 45692,8. El tamaño de la muestra es de 52 datos.

En la figura 4.39 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 32215,85 y de igual

manera en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 40438,5. En ésta variable existen valores extremos.

Figura 4.39 Diagramas de Cajas THSW Index Acumulada



Con los datos obtenidos de asimetría $-0,46$ decimos que la curva es *simétrica* ya que se encuentra entre (± 0.5) , es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media (véase figura 4.40), en cuanto a la puntiagudez respecto a la normal la distribución es *mesocúrtica* ya que el coeficiente de curtosis es negativo con un valor de $-0,31$ y encontrándose entre (± 0.5) .

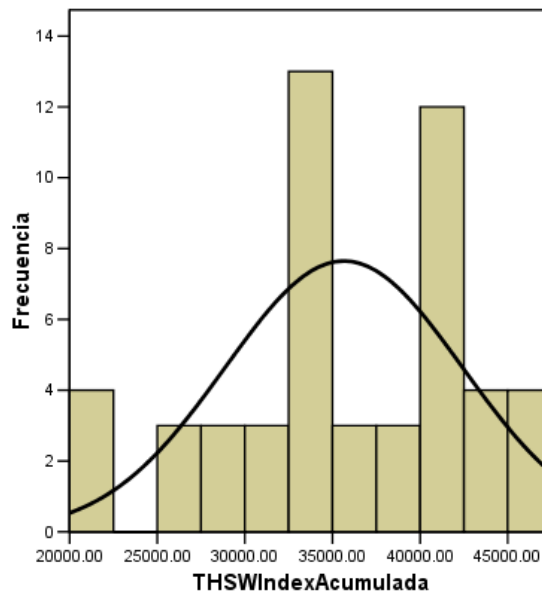


Figura 4.40 Histograma de THSW Index Acumulada

4.3.3.16. THSW INDEX PROMEDIO

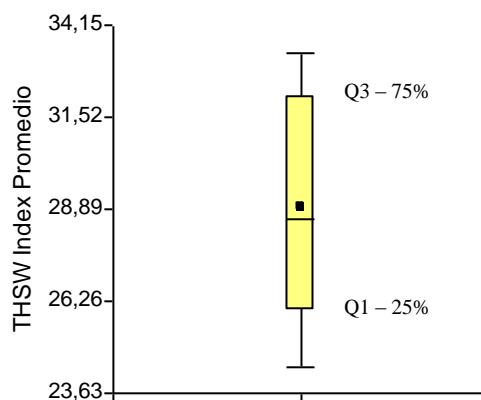
La variable THSW Index Acumulada es el índice que relaciona la temperatura, humedad, radiación solar y viento promediado y expresado ésta en grados centígrados (°C).

Se obtuvo una media de 28,96 con una desviación estándar de 3,27 un valor mínimo de 24,11 y un valor máximo de 33,67. El tamaño de la muestra es de 52 datos.

En la figura 4.41 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 25,97 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 32,1. En ésta variable no existen valores extremos.

Figura 4.41 Diagrama de Cajas de THSW Index Promedio



Con los datos obtenidos de asimetría 0,03 decimos que la curva es *simétrica* ya que se encuentra entre (± 0.5) , es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media (véase figura 4.42), en cuanto a la puntigudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es negativo con un valor de -1,42.

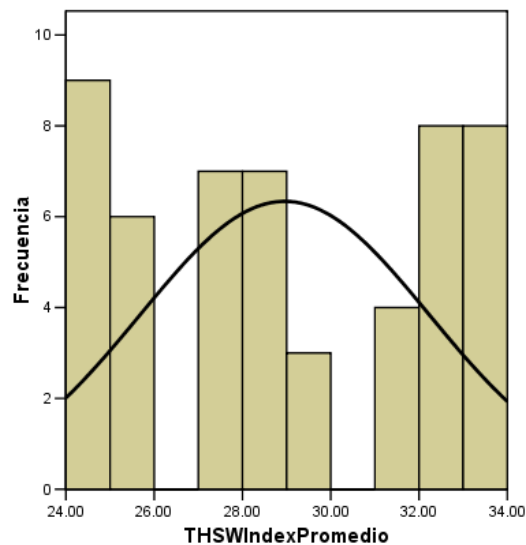


Figura 4.42 Histograma de THSW Index Promedio

4.3.3.17. COOL D-D ACUMULADA

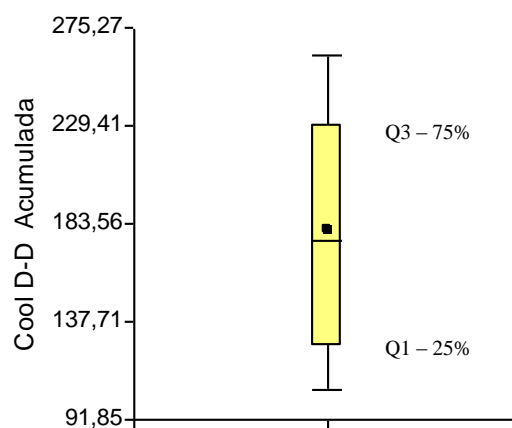
La variable Cool D-D Acumulada se obtiene del promedio de los registros climáticos acumulados mensualmente.

Se obtuvo una media de 180,44 con una desviación estándar de 58,92 un valor mínimo de 100,19 y un valor máximo de 266,93. El tamaño de la muestra es de 52 datos.

En la figura 4.43 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 125,97 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 229,45. En ésta variable no existen valores extremos.

Figura 4.43 Diagrama de Cajas de Cool D-D Acumulada



Con los datos obtenidos de asimetría 0,08 decimos que la curva es simétrica ya que se encuentra entre (± 0.5) (véase figura 4.44) es decir, existe aproximadamente la misma cantidad de valores a los dos lados de la media, en cuanto a la puntiagudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es -1,49.

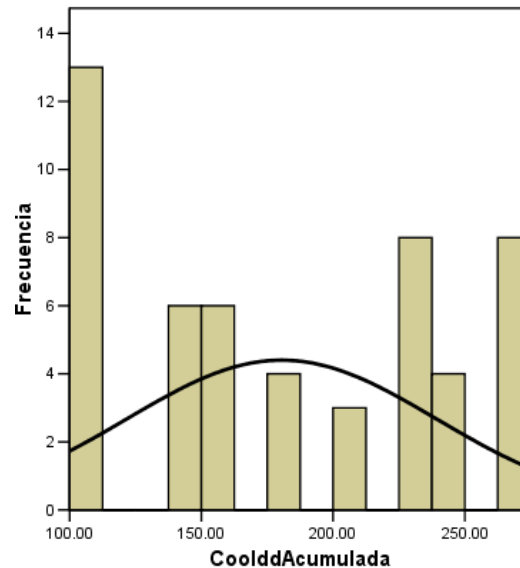


Figura 4.44 Histograma Cool D-D Acumulada

4.3.3.18. COOL D-D PROMEDIO

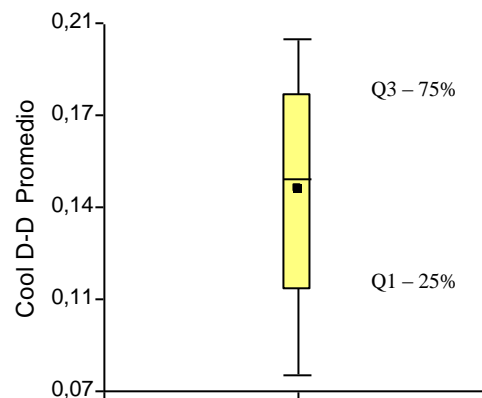
La variable Cool D-D Promedio se obtiene del promedio de los registros climáticos

Se obtuvo una media de 0,15 con una desviación estándar de 0,04 un valor mínimo de 0,08 y un valor máximo de 0,2. El tamaño de la muestra es de 52 datos.

En la figura 4.45 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 0,11 y de igual manera en

el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 01,8. En ésta variable no existen valores extremos.

Figura 4.45 Diagrama de Cajas de Cool D-D Promedio



Con los datos obtenidos de asimetría $-0,21$ decimos que la curva es simétrica ya que se encuentra entre (± 0.5) (véase figura 4.46) es decir, existe aproximadamente la misma cantidad de valores a los dos lados de la media, en cuanto a la puntigudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es $-1,29$

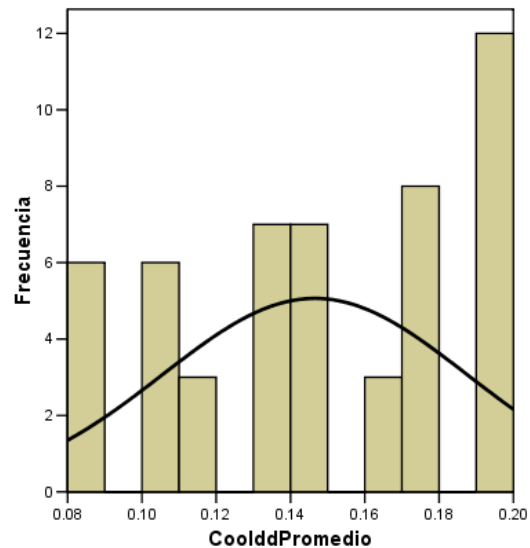


Figura 4.46 Histograma de Cool D-D Promedio

4.3.3.19. SOLAR RAD ACUMULADA

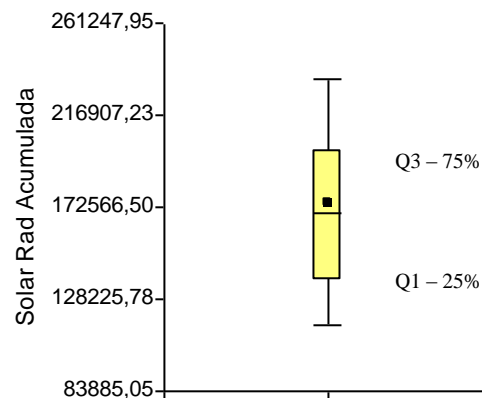
La variable Solar Rad Acumulada es la radiación solar registrada en w / m^2 . Acumulado por mes.

Se obtuvo una media de 174927,48 con una desviación estándar de 45634,5 un valor mínimo de 91947 y un valor máximo de 253186. El tamaño de la muestra es de 52 datos.

En la figura 4.47 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 137172 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 199199. En ésta variable no existen valores extremos.

Figura 4.47 Diagrama de Cajas de Solar Rad Acumulada



Con los datos obtenidos de asimetría 0,03 decimos que la curva es simétrica ya que se encuentra entre (± 0.5) (véase figura 4.48), es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media, en cuanto a la puntiagudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es -0,72.

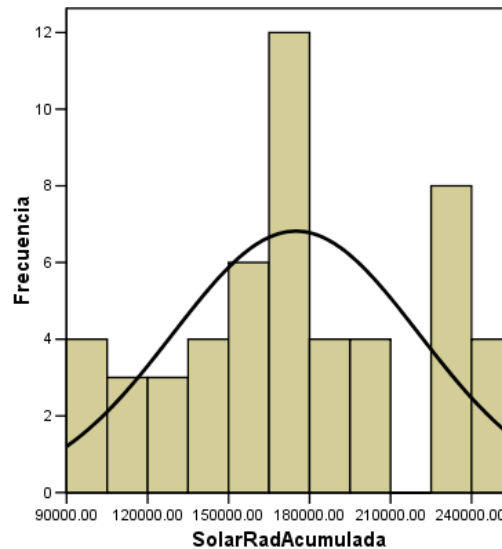


Figura 4.48 Histograma de Solar Rad Acumulada

4.3.3.20. SOLAR RAD PROMEDIO

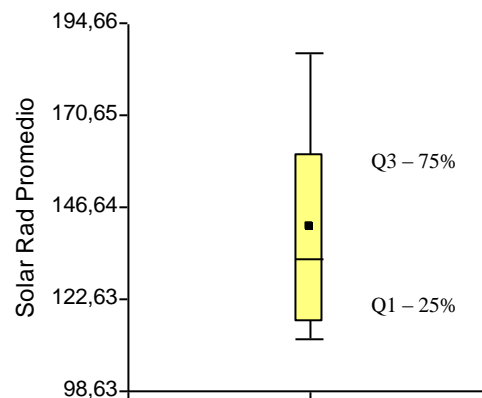
La variable Solar Rad Promedio es la radiación solar promedio mes registrada en w / m^2 .

Se obtuvo una media de 141,54 con una desviación estándar de 27,9 un valor mínimo de 102,99 y un valor máximo de 190,29. El tamaño de la muestra es de 52 datos.

En la figura 4.49 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 116,15 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 160. En ésta variable no existen valores extremos.

Figura 4.49 Diagrama de Cajas Solar Rad Promedio



Con los datos obtenidos de asimetría 0,44 decimos que la curva es simétrica ya que se encuentra entre (± 0.5) (véase figura 4.50), es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media, en cuanto a la puntiagudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es -1,14.

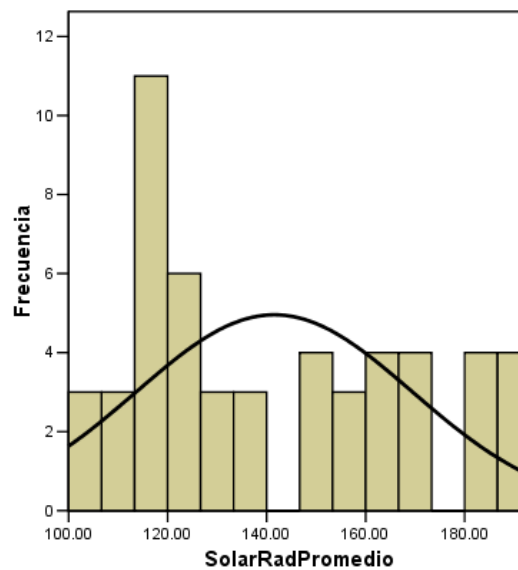


Figura 4.50 Histograma de Solar Rad Promedio

4.3.3.21. SOLAR ENERGY ACUMULADA

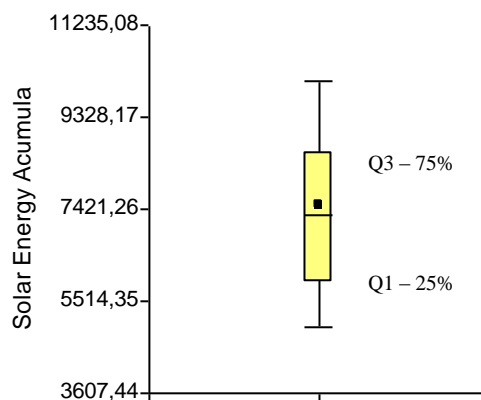
La variable Solar Energy Acumulada es la energía solar acumulada por m^2 registrada en Langley.

Se obtuvo una media de 7522,86 con una desviación estándar de 1962,57 un valor mínimo de 3954,14 y un valor máximo de 10888,37. El tamaño de la muestra es de 52 datos.

En la figura 4.51 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 5899,13 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 8566,79. En ésta variable no existen valores extremos.

Figura 4.51 Diagrama de Cajas Solar Energy Acumula



Con los datos obtenidos de asimetría 0,03 decimos que la curva es simétrica ya que se encuentra entre (± 0.5) (véase figura 4.52), es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media, en cuanto a la puntiagudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es -0,72.

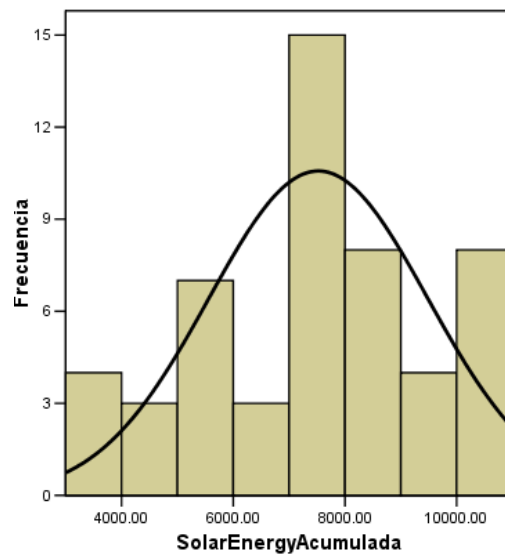


Figura 4.52 Histograma de Solar Energy Acumulada

4.3.3.22. SOLAR ENERGY PROMEDIO

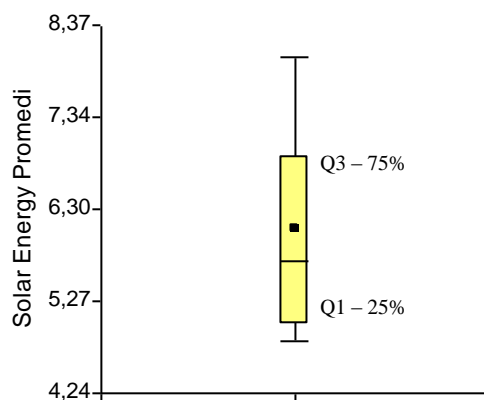
La variable Solar Energy Promedio es la energía solar promediada por m^2 registrada en Langley.

Se obtuvo una media de 6,09 con una desviación estándar de 1,2 un valor mínimo de 4,43 y un valor máximo de 8,18. El tamaño de la muestra es de 52 datos.

En la figura 4.53 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 5 y de igual manera en el

cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 6,88. En ésta variable no existen valores extremos.

Figura 4.53 Diagrama de Cajas de Solar Energy Promedi



Con los datos obtenidos de asimetría 0,43 decimos que la curva es simétrica ya que se encuentra entre (± 0.5) (véase figura 4.54), es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media, en cuanto a la puntiagudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es -1,14.

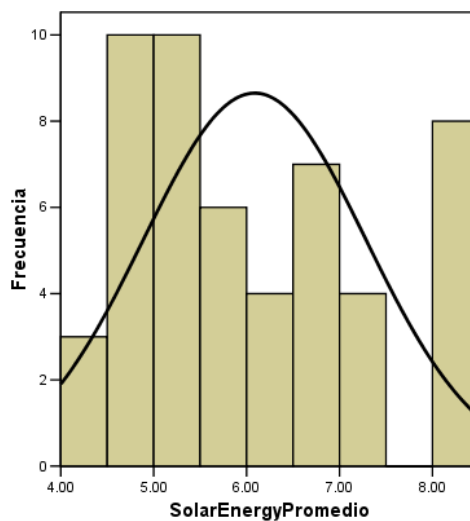


Figura 4.54 Histograma Solar Energy Promedio

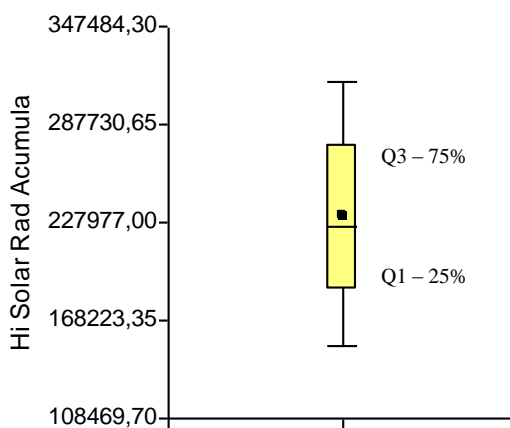
4.3.3.23. HI SOLAR RAD ACUMULADA

La variable Hi Solar Rad Acumulada es la radiación máxima registrada en Langley.

Se obtuvo una media de 231761,37 con una desviación estándar de 61411,93 un valor mínimo de 119334 y un valor máximo de 335520. El tamaño de la muestra es de 52 datos.

En la figura 4.55 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 185917 y de igual manera en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 274208. En ésta variable no existen valores extremos.

Figura 4.55 Diagramas de Cajas de Hi Solar Rad Acumula



Con los datos obtenidos de asimetría 0,04 decimos que la curva es simétrica ya que se encuentra entre (± 0.5) (véase figura 4.56), es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media, en cuanto a la puntiagudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es -0,7.

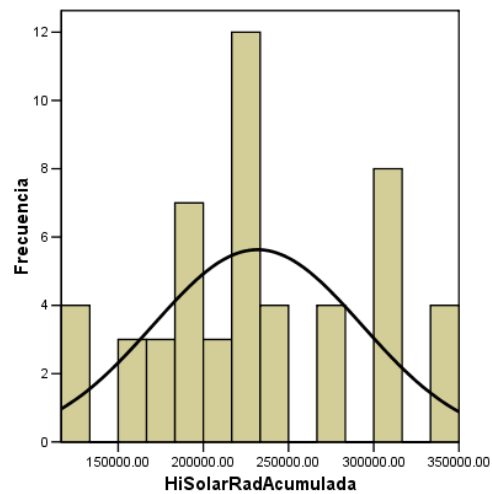


Figura 4.56 Histograma de Hi Solar Rad Acumulada

4.3.3.24. HI SOLAR RAD PROMEDIO

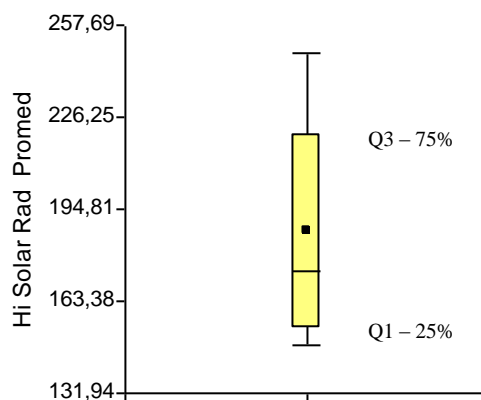
La variable Hi Solar Rad Promedio es la radiación solar máxima registrada en Langley y promediada ésta por mes.

Se obtuvo una media de 187,49 con una desviación estándar de 38,06 un valor mínimo de 137,66 y un valor máximo de 251,97. El tamaño de la muestra es de 52 datos.

En la figura 4.57 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 154,03 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 220,25. En ésta variable no existen valores extremos.

Figura 4.57 Diagramas de Cajas Hi Solar Rad Promedio



Con los datos obtenidos de asimetría 0,46 decimos que la curva es simétrica ya que se encuentra entre (± 0.5) (véase figura 4.58), es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media, en cuanto a la puntiagudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es -1,28.

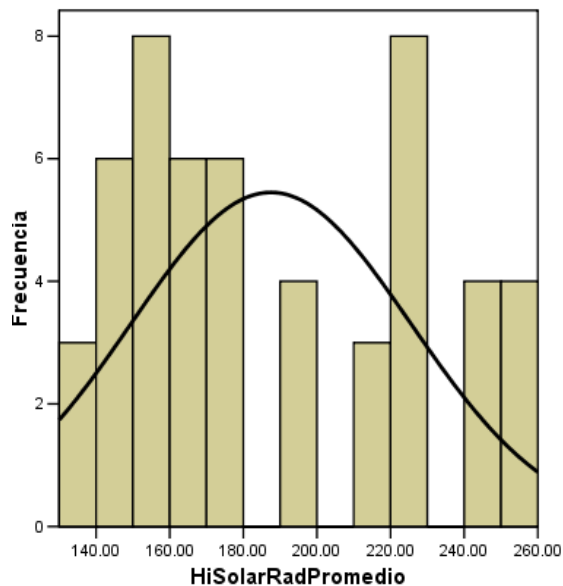


Figura 58 Histograma de Hi Solar Rad Promedio

4.3.3.25. UV INDEX ACUMULADA

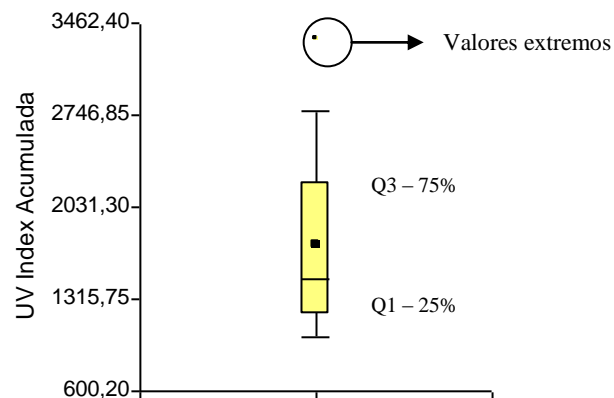
La variable UV Index Acumulada es la periodicidad de rayos ultravioletas registrada en la escala de [1-16] y acumulada por mes.

Se obtuvo una media de 1742,51 con una desviación estándar de 727,51 un valor mínimo de 730,3 y un valor máximo de 3332,3. El tamaño de la muestra es de 52 datos.

En la figura 8.59 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 1193,4 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 2212,80. En ésta variable existen valores extremos.

Figura 4.59 Diagramas de Cajas UV Index Acumulada



Con los datos obtenidos de asimetría 0,76 decimos que la curva es asimétrica positiva por lo que los valores se tienden a reunir más en la parte izquierda que en la derecha de la media (véase figura 4.60), en cuanto a la puntiagudez respecto a la normal la distribución es *mesocúrtica* ya que el coeficiente de curtosis es aproximado a cero $\pm 0,5$ con un valor de -0,27.

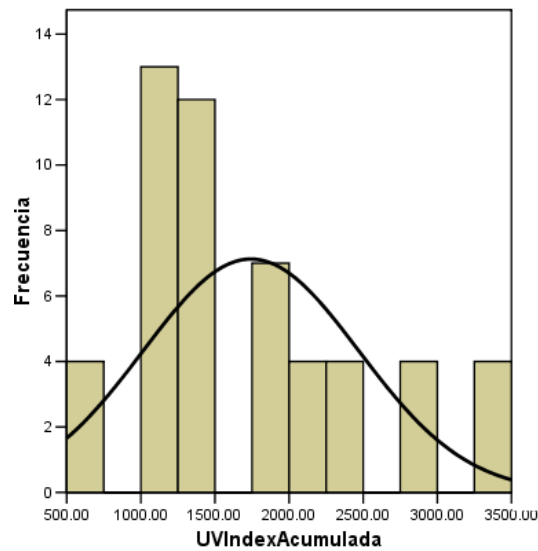


Figura 4.60 Histograma de UV Index Acumulada

4.3.3.26. UV INDEX PROMEDIO

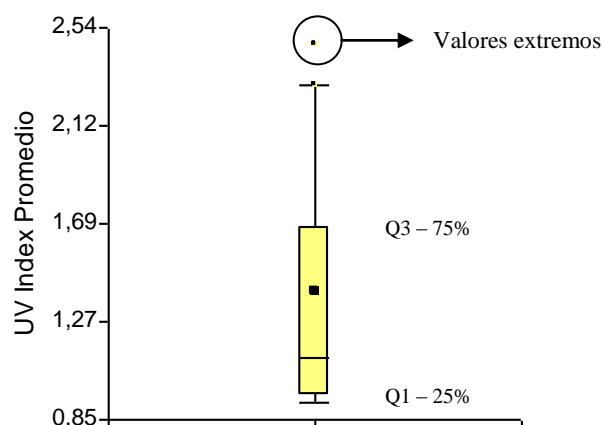
La variable UV Index Promedio es la periodicidad de rayos ultravioletas promedio registrada en escala de [1-16] mensual.

Se obtuvo una media de 1,4 con una desviación estándar de 0,52 un valor mínimo de 0,93 y un valor máximo de 2,46. El tamaño de la muestra es de 52 datos.

En la figura 4.61 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 0,95 y de igual manera en

el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 1,68. En ésta variable existen valores extremos.

Figura 4.61 Diagrama de Cajas de UV Index Promedio



Con los datos obtenidos de asimetría 0,87 decimos que la curva es asimétrica positiva por que los valores tienden a reunir más en la parte izquierda que en la derecha de la media (véase figura 4.62), en cuanto a la puntiagudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es -0,6.

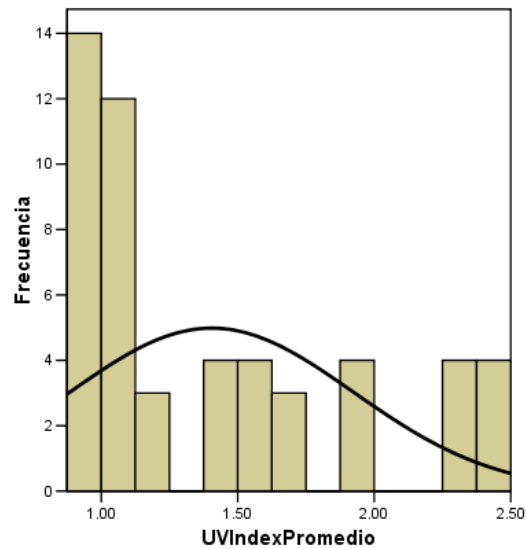


Figura 4.62 Histograma de UV Index Promedio

4.3.3.27. UV DOSE ACUMULADA

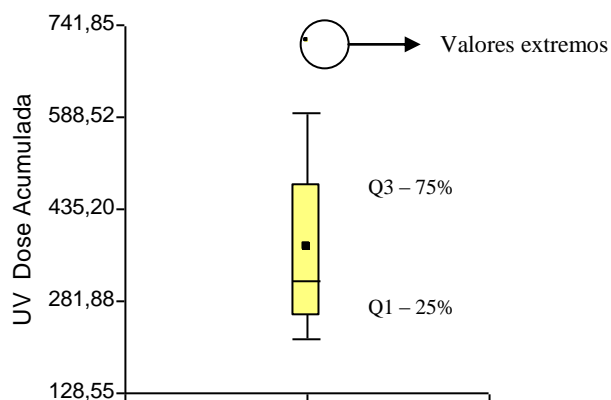
La variable UV Dose Acumulada es la dosis de radiación ultravioleta acumulada mensualmente.

Se obtuvo una media de 373,34 con una desviación estándar de 155,9 un valor mínimo de 156,43 y un valor máximo de 713,97. El tamaño de la muestra es de 52 datos.

En la figura 4.63 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 255,73 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 474,1. En ésta variable existen valores extremos.

Figura 4.63 Diagrama de Cajas de UV Dose Acumulada



Con los datos obtenidos de asimetría 0,76 decimos que la curva es asimétrica positiva por que los valores se tienden a reunir más en la parte izquierda que en la derecha de la media (véase figura 4.64), en cuanto a la puntiagudez respecto a la normal la distribución es *mesocúrtica* ya que el coeficiente de curtosis es aproximado a cero $\pm 0,5$ con un valor de -0,27.

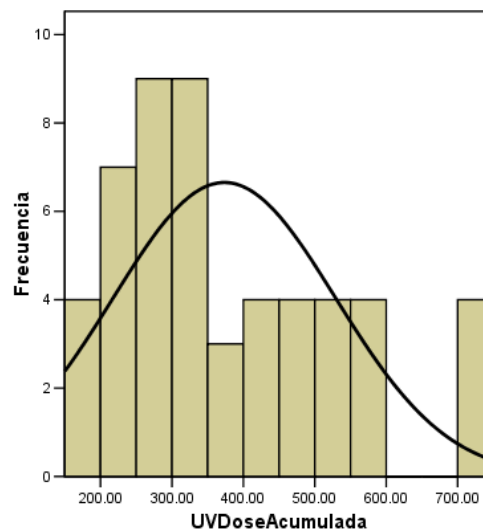


Figura 4.64 Histograma de UV Dose Acumulada

4.3.3.28. UV DOSE PROMEDIO

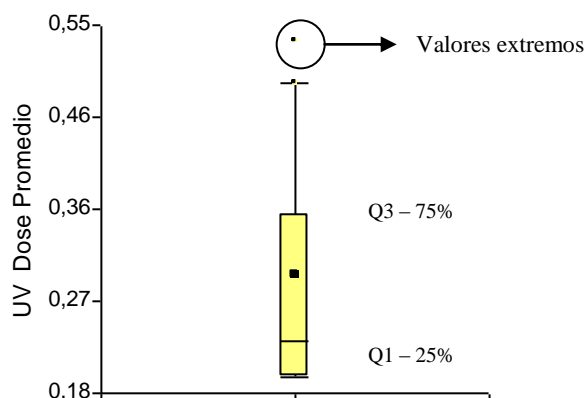
La variable UV Dose Promedio es la dosis de radiación ultravioleta promedio mensual.

Se obtuvo una media de 0,3 con una desviación estándar de 0,11 un valor mínimo de 0,2 y un valor máximo de 0,53. El tamaño de la muestra es de 52 datos.

En la figura 4.65 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 0,2 y de igual manera en

el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 0,36. En ésta variable existen valores extremos.

Figura 4.65 Diagrama de Cajas de UV Dose Promedio



Con los datos obtenidos de asimetría 0,87 decimos que la curva es asimétrica positiva por que los valores se tienden a reunir más en la parte izquierda que en la derecha de la media (véase figura 4.66), en cuanto a la puntiagudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es -0,62.

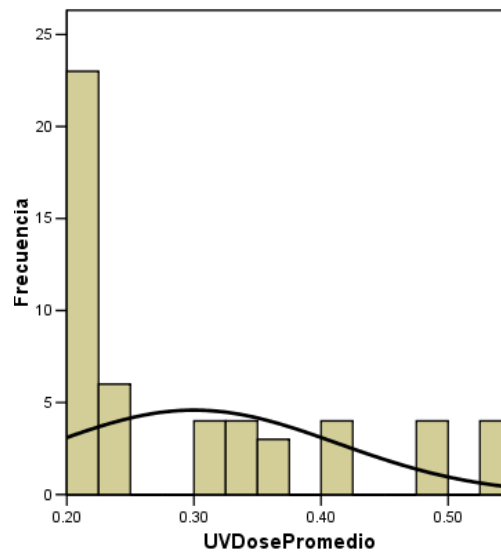


Figura 4.66 Histograma de UV Dose Promedio

4.3.3.29. HI UV ACUMULADA

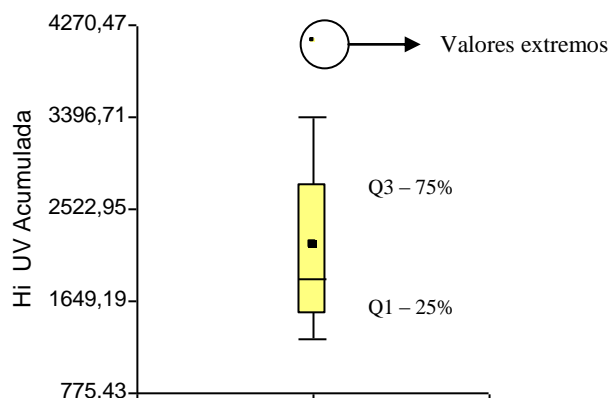
La variable Hi UV Acumulada es el índice máximo de rayos ultravioletas acumulados y registrados mensualmente.

Se obtuvo una media de 2187,31 con una desviación estándar de 887,7 un valor mínimo de 934,3 y un valor máximo de 4111,6. El tamaño de la muestra es de 52 datos.

En la figura 4.67 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 1524,5 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 2760,2. En ésta variable existen valores extremos.

Figura 4.67 Diagrama de Cajas de Hi UV Acumulada



Con los datos obtenidos de asimetría 0,74 decimos que la curva es asimétricamente positiva por que los valores se tienden a reunir más en la parte izquierda que en la derecha de la media (véase figura 4.68), en cuanto a la puntiagudez respecto a la normal la distribución es *mesocúrtica* ya que el coeficiente de curtosis es aproximado a cero $\pm 0,5$ con un valor de -0,33.

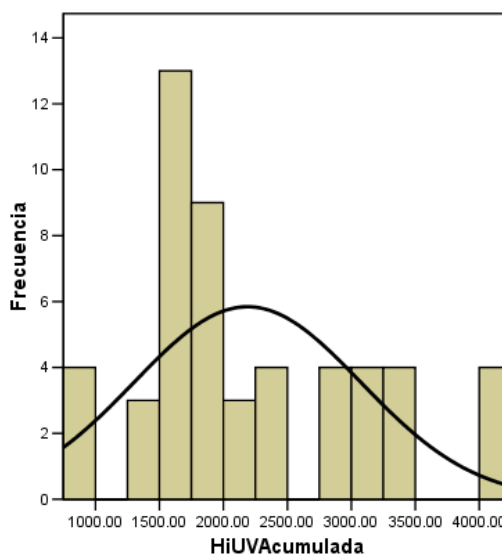


Figura 4.68 Histograma de Hi UV Acumulada

4.3.3.30. HI UV PROMEDIO

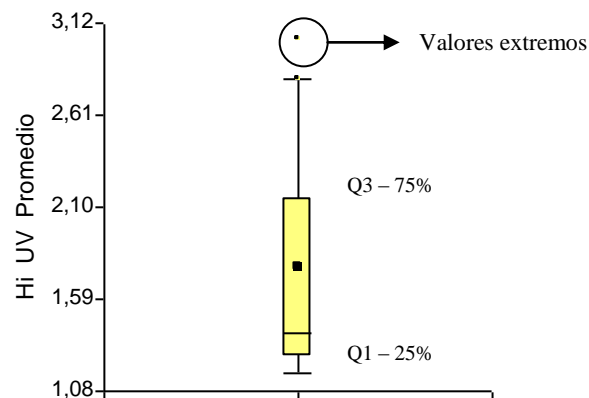
La variable Hi UV Promedio es el índice máximo de rayos ultravioletas.

Se obtuvo una media de 1,76 con una desviación estándar de 0,63 un valor mínimo de 1,17 y un valor máximo de 3,03. El tamaño de la muestra es de 52 datos.

En la figura 4.69 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 1,27 y de igual manera en

el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 2,14. En ésta variable existen valores extremos.

Figura 4.69 Diagrama de Cajas de Hi UV Promedio



Con los datos obtenidos de asimetría 0,85 decimos que la curva es asimétricamente positiva por que los valores se tienden a reunir más en la parte izquierda que en la derecha de la media (véase figura 4.70), en cuanto a la puntiagudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es -0,73.

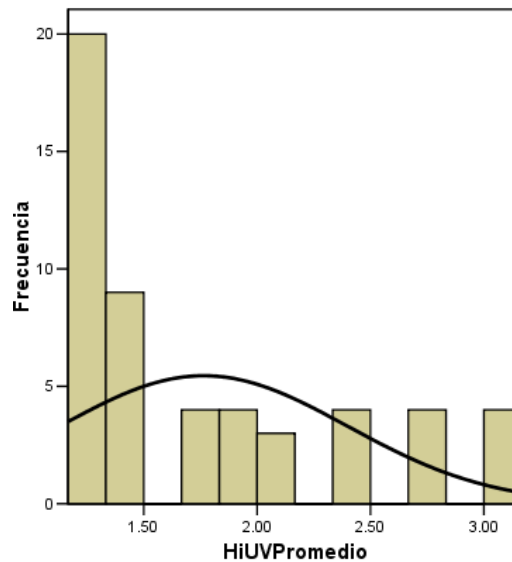


Figura 4.70 Histograma de Hi UV Promedio

4.3.3.31. ET ACUMULADA

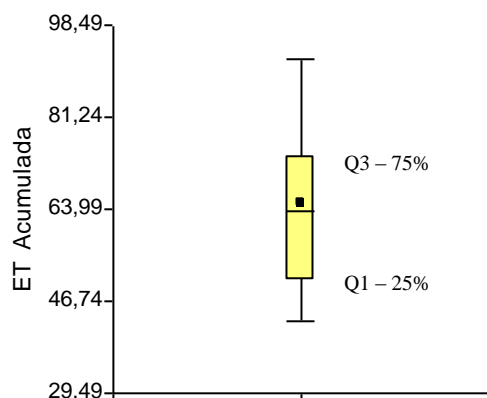
La variable ET Acumulada es la evotranspiración registrada en mm acumulada mensualmente.

Se obtuvo una media de 65,15 con una desviación estándar de 18,1 un valor mínimo de 32,63 y un valor máximo de 95,35. El tamaño de la muestra es de 52 datos.

En la figura 4.71 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 50,61 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 73,74. En ésta variable no existen valores extremos.

Figura 4.71 Diagramas de Cajas de ET Acumulada



Con los datos obtenidos de asimetría 0,12 decimos que la curva es simétrica ya que se encuentra entre $\pm 0,5$ (véase figura 4.72), es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media, en cuanto a la puntiagudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es -0,7.

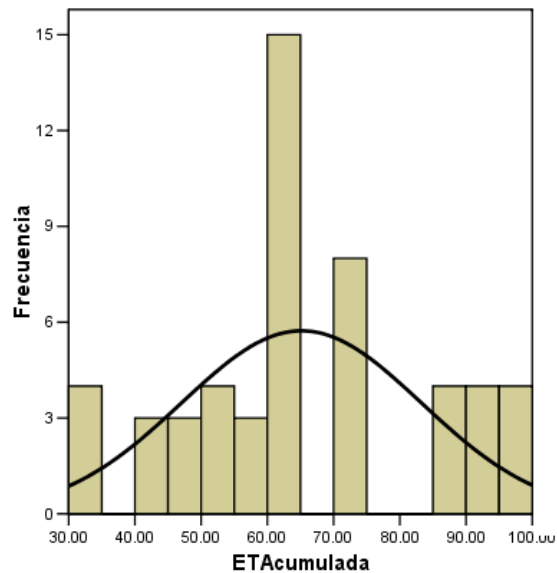


Figura 4.72 Histograma de ET Acumulada

4.3.3.32. ET PROMEDIO

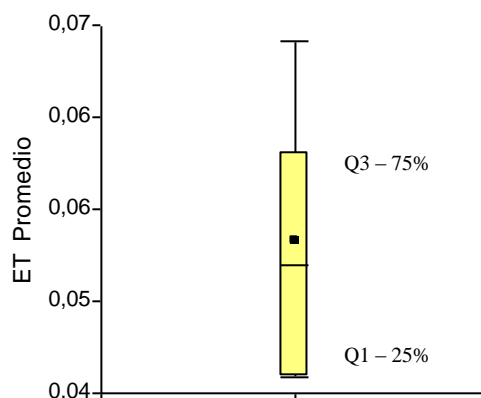
La variable Et Promedio es la evotranspiración promedio y registrado en mm por mes.

Se obtuvo una media de 0,05 con una desviación estándar de 0,01 un valor mínimo de 0,04 y un valor máximo de 0,07. El tamaño de la muestra es de 52 datos.

En la figura 4.73 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 0,04 y de igual manera en

el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 0,06. En ésta variable no existen valores extremos.

Figura 4.73 Diagrama de Cajas de ET Promedio



Con los datos obtenidos de asimetría 0,44 decimos que la curva es simétrica ya que se encuentra entre $\pm 0,5$ (véase figura 4.74), es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media, en cuanto a la puntiagudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es -1,35.

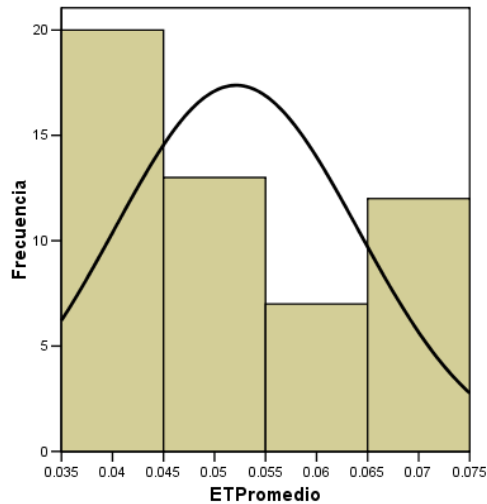


Figura 4.74 Histograma de ET Promedio

4.3.3.33. OUT HUM ACUMULADA

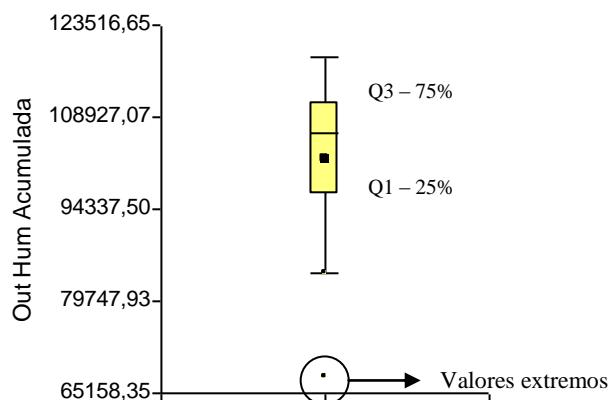
La variable Out Hum Acumulada es la humedad relativa exterior expresada en porcentaje acumulada mensualmente.

Se obtuvo una media de 102348,96 con una desviación estándar de 13942,33 un valor mínimo de 67811 y un valor máximo de 120864. El tamaño de la muestra es de 52 datos.

En la figura 4.75 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 96608 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 111110. En ésta variable existen valores extremos.

Figura 4.75 Diagramas de Cajas de Out Hum Acumulada



Con los datos obtenidos de asimetría $-1,18$ decimos que la curva es asimétricamente negativa por que los valores se tienden a reunir más en la parte derecha de la media (véase figura 4.76), en cuanto a la puntiagudez respecto a la normal la distribución es *leptocúrtica* ya que el coeficiente de curtosis es $0,88$.

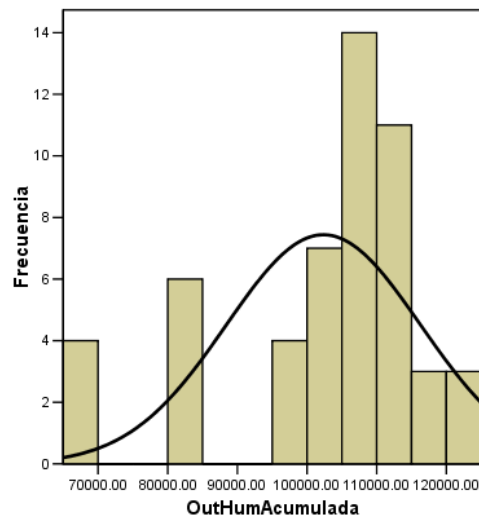


Figura 4.76 Histograma de Out Hum Acumulada

4.3.3.34. OUT HUM PROMEDIO

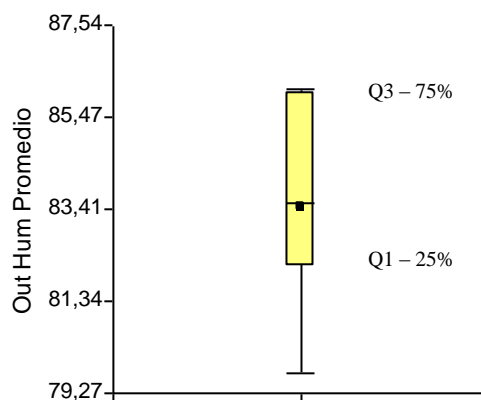
La variable Out Hum Promedio es la Humedad Relativa Exterior promedio expresada en porcentaje mensual.

Se obtuvo una media de 83,44 con una desviación estándar de 2,41 un valor mínimo de 79,65 y un valor máximo de 87,16. El tamaño de la muestra es de 52 datos.

En la figura 4.77 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 82,12 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 86,03. En ésta variable no existen valores extremos.

Figura 4.77 Diagramas de Cajas de Out Hum Promedio



Con los datos obtenidos de asimetría $-0,2$ decimos que la curva es simétrica ya que se encuentra entre $\pm 0,5$ (véase figura 4.78), es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media, en cuanto a la puntigudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es $-1,02$.

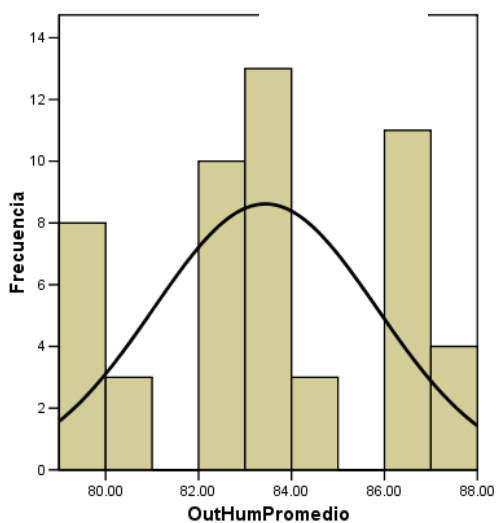


Figura 78 Histograma de Out Hum Promedio

4.3.3.35. DEW PT. ACUMULADA

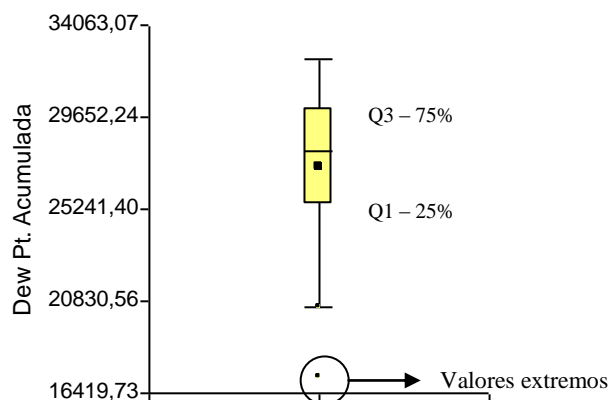
La variable Dew Pt. Acumulada es el punto de condensación acumulada mensualmente y expresada ésta en grados centígrados (°C).

Se obtuvo una media de 27236,12 con una desviación estándar de 4375,22 un valor mínimo de 17221,7 y un valor máximo de 33261,1. El tamaño de la muestra es de 52 datos.

En la figura 4.79 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 25470,35 y de igual

manera en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 30078,4. En ésta variable no existen valores extremos.

Figura 4.79 Diagrama de Cajas de Dew Pt. Acumulada



Con los datos obtenidos de asimetría $-0,87$ decimos que la curva es asimétricamente negativa por que los valores se tienden a reunir más en la parte derecha de la media (véase figura 4.80), en cuanto a la puntiagudez respecto a la normal la distribución es *mesocúrtica* ya que el coeficiente de curtosis es aproximado a cero $\pm 0,5$ con un valor de $0,23$.

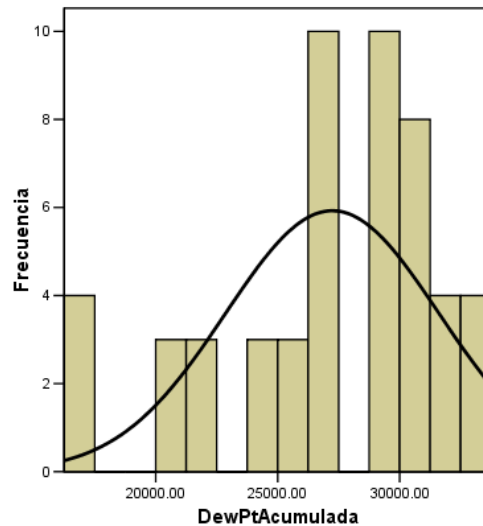


Figura 4.80 Histograma de Dew Pt. Acumulada

4.3.3.36. DEW PT. PROMEDIO

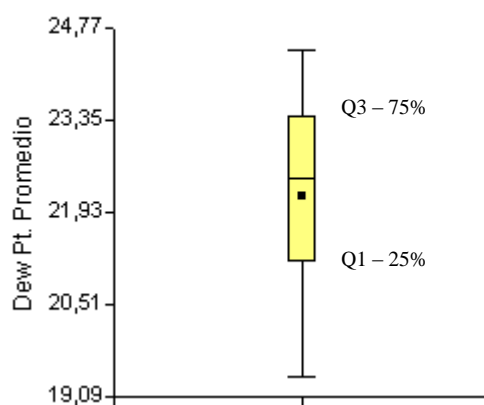
La variable Dew Pt. Promedio es el punto de condensación promedio mensual expresada éste en grados centígrados (°C).

Se obtuvo una media de 22,17 con una desviación estándar de 1,7 un valor mínimo de 19,35 y un valor máximo de 24,51. El tamaño de la muestra es de 52 datos.

En la figura 4.81 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 21,16 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 23,41. En ésta variable no existen valores extremos.

Figura 4.81 Diagrama de Cajas de Dew Pt. Promedio



Con los datos obtenidos de asimetría $-0,31$ decimos que la curva es simétrica ya que se encuentra entre $\pm 0,5$ (véase figura 4.82), es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media, en cuanto a la puntiagudez respecto a la normal la distribución es *platicúrtica* ya que el coeficiente de curtosis es $-1,1$.

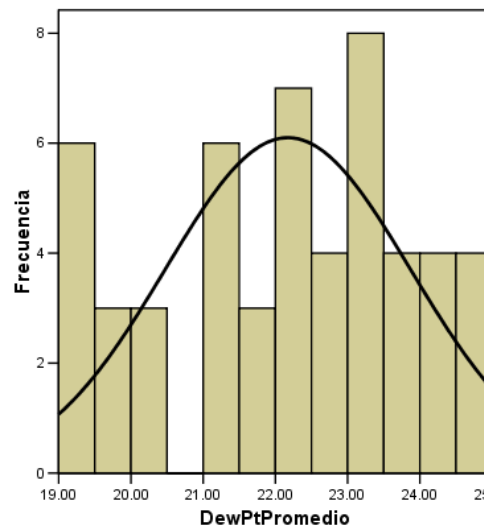


Figura 4.82 Histograma de Dew Pt. Promedio

4.3.3.37. WIND SPEED ACUMULADA

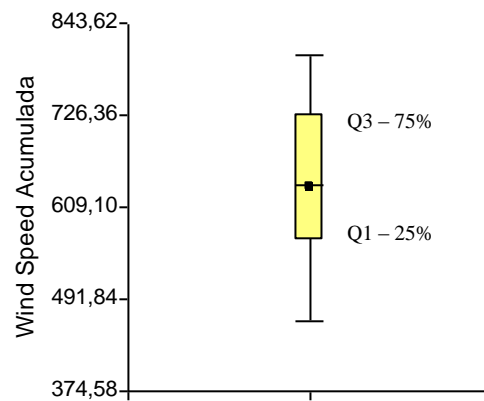
La variable Wind Speed Acumulada es la velocidad del viento acumulada mensualmente.

Se obtuvo una media de 634,33 con una desviación estándar de 116,78 un valor mínimo de 395,9 y un valor máximo de 822,3. El tamaño de la muestra es de 52 datos.

En la figura 4.83 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 566,3 y de igual manera

en el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 727,4. En ésta variable no existen valores extremos.

Figura 4.83 Diagrama de Cajas de Wind Speed Acumulada



Con los datos obtenidos de asimetría $-0,26$ decimos que la curva es asimétricamente negativa por que los valores se tienden a reunir más en la parte derecha de la media (véase figura 4.84), en cuanto a la puntiagudez respecto a la normal la distribución es *mesocúrtica* ya que el coeficiente de curtosis es aproximado a cero $\pm 0,5$ con un valor de $-0,28$.

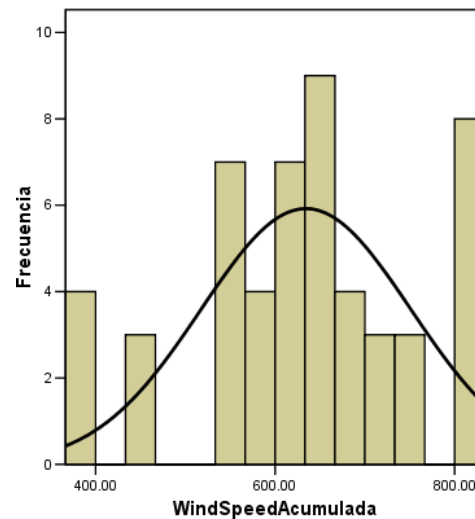


Figura 4.84 Histograma de Wind Speed Acumulada

4.3.3.38. WIND SPEED PROMEDIO

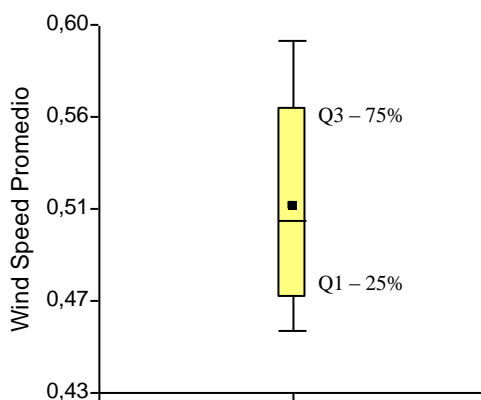
La variable Wind Speed Promedio es la velocidad del viento promediada mensualmente.

Se obtuvo una media de 0,52 con una desviación estándar de 0,05 un valor mínimo de 0,44 y un valor máximo de 0,59. El tamaño de la muestra es de 52 datos.

En la figura 4.85 determinamos que en el cuartil 1 (Q_1) el 25% de las observaciones toman valores menores o iguales a 0,48 y de igual manera en

el cuartil 3 (Q_3) el 75% de las observaciones toman valores menores o iguales a 0,56. En ésta variable no existen valores extremos.

Figura 4.85 Diagramas de Cajas de Wind Speed Promedio



Con los datos obtenidos de asimetría 0,27 decimos que la curva es simétrica ya que se encuentra entre $\pm 0,5$ (véase figura 4.86), es decir existe aproximadamente la misma cantidad de valores a los dos lados de la media, en cuanto a la puntiagudez respecto a la normal la distribución es *leptocúrtica* ya que el coeficiente de curtosis es -1,22.

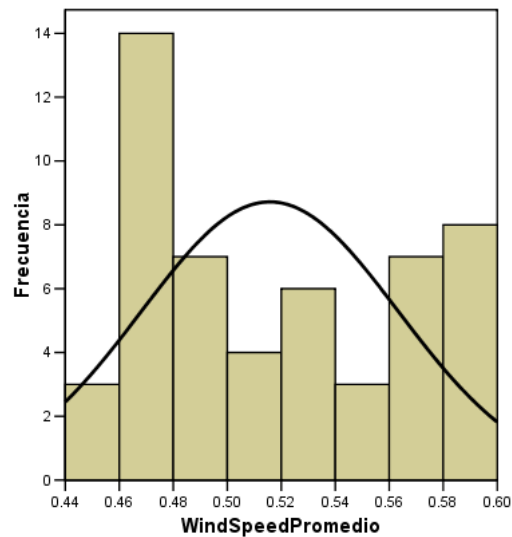


Figura 4.86 Histograma de Wind Speed Promedio

4.4. BONDAD DE AJUSTE (KOLMOGOROV SMIRNOV)

Con la Prueba de Bondad de Ajuste de Kolmogorov Smirnov vamos a probar si la distribución empírica de la variable aleatoria x se ajusta a una distribución *Normal* (μ, σ^2) , con un nivel de significancia de 1%.

La prueba de hipótesis es:

$$H_0: G(x) = \text{Normal}(\mu, \sigma^2)$$

vs

$$H_1: G(x) \neq \text{Normal}(\mu, \sigma^2) \text{ para al menos una } x$$

En el anexo IV, podemos observar las variables con su media, desviación estándar, el valor del estadístico de Kolmogorov Smirnov y su respectivo valor p .

De todas las variables describiremos las que tengan valores de p menores al nivel de significancia, ya que en éstas se rechaza la Hipótesis Nula o H_0 ; es decir, se acepta que la distribución empírica es diferente de la distribución normal o H_1 .

Éste grupo de variables son (véase anexo IV): Rain Acumulada, Rain Promedio, UV Index Promedio, UV Dose Promedio, Hi UV Acumulada, Hi UV Promedio, ET Promedio, Out Hum Acumulada.

Tiene relación las variables que rechazan la hipótesis nula ya que Rain es lluvia, UV son los rayos ultravioletas y el Out Hum es la humedad relativa exterior, ya que tienen una periodicidad anual, presentan valores de asimetría considerables como es el caso de Rain con una asimetría de 2,11, Rayos Ultravioletas con una asimetría 0,86, y la Humedad Externa con una asimetría de -1,18.

4.5. ANÁLISIS DE CORRELACIÓN (DE PEARSON)

Con ésta información podemos analizar las variables que tienen una mayor relación entre las variables de clima con las variables de preaviso y severidad.

Determinando la correlación de Pearson entre todas las variable de clima e infección, con un nivel de significancia del 5%, los datos pueden observarse en anexo I donde el valor de p es la segunda fila de cada variable tomando en cuenta que si $p < 0.05$ la relación entre variables es más fuerte, las variables más relacionadas entre sí son: III, IV, EE, Floración, Cosecha, Rain Acumulada (Precipitación), Temp Out Promedio, Temp Hi Promedio, Temp Low Promedio, Wind Chill Promedio, Heat Index Promedio, THW Index Promedio, THSW Index Promedio, Cool D-D Promedio, Solar Rad Promedio, Solar Energy Promedio, Hi Solar Rad Promedio, UV Index Promedio, UV Dose Promedio, Hi UV Promedio, Dew Pt. Promedio.

Analizando las variables de preaviso III, IV podemos observar en el figura 87, que los datos tienen la misma tendencia, por tal razón escogeremos una de las variables en éste caso la variable III.

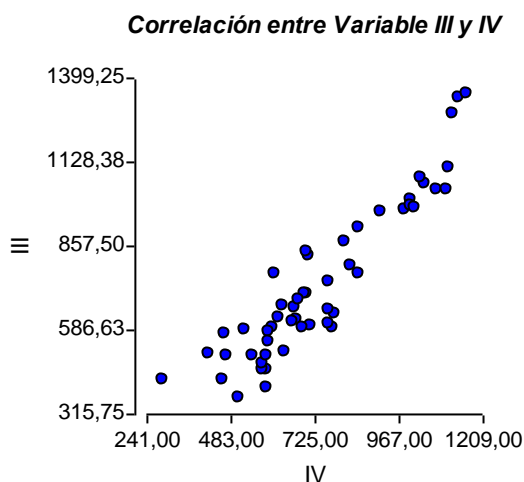


Figura 4.87

Así mismo observando la correlación entre las variables de temperatura la que más relación tiene con tres o más variables de preaviso y severidad es Temp Low Promedio (Temperatura mínima exterior en °C), así mismo tenemos las variables que indican el índice que relaciona la humedad, temperatura y viento en éste caso la que tiene una mayor relación es THW Promedio, en el caso de las variables relacionadas con la energía solar tenemos Solar Rad Promedio (Radiación solar registrada en w / m^2), entre las variables relacionadas con los rayos ultravioletas la que más relación tiene es UV Dose Promedio (Dosis de Radiación Ultravioleta).

Entonces las variables con las que vamos a desarrollar nuestro análisis Multivariado son: Rain Acumulada, Temp Low Promedio, THW Index Promedio, Cool D-D Promedio, Solar Rad Promedio, UV Dose Promedio, ET

Promedio, Dew Pt. Promedio, ya que estas son las que proveen de más información para el análisis, una vez observados los resultados del anexo I.

4.6. ANÁLISIS DE CONGLOMERADOS

El objetivo del análisis de conglomerados (*o cluster análisis*) es la clasificación (no supervisada) de los elementos de ε , es decir, su agrupación en clases disjuntas, que se denominan conglomerados (*o clusters*).

En el anexo V tenemos el vector de medias que nos servirán para el análisis de las variables de infección en éste caso las variables III, EE, Floración y Cosecha en los años 2005 y 2006, el criterio de clasificación es por lotes en éste caso A05, A06, B05, B06, C05, C06, D06 que significa el lote A del año 2005, el lote B del año 2005 y así sucesivamente.

En la figura 4.88 podemos observar el dendograma de análisis de conglomerados de las variables de infección, con una correlación cofenética de 0,891; se debe recordar que mientras más se aproxime a 1 el valor de la correlación cofenética, es mejor el agrupamiento.

Análisis de Conglomerados Variables Infección

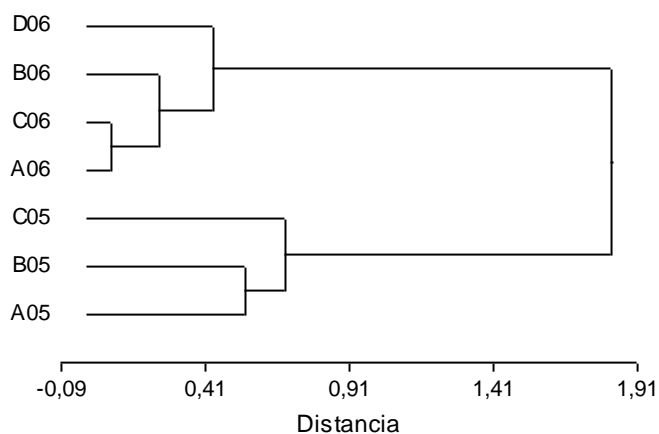


Figura 4.88

Observando el dendrograma los lotes A06, C06, B06 y D06 forman un solo grupo y así los lotes A05, B05 y C05 forman otro grupo. Lo que nos indica que los lotes en el año 2005 forman un solo grupo. Así también podemos agrupar los lotes del 2006. La similitud de los miembros dentro de cada grupo, es mayor al 90%.

nodo1	nodo2	Distancia
A06	C06	0,08
A06	B06	0,25
A06	D06	0,43
A05	B05	0,55
A05	C05	0,68
A05	A06	1,82

Tabla I Distancia entre Grupos

En la tabla I podemos observar las distancias entre los dos grupos. En el lote A06 y C06 la distancia es de 0,08, en el lote A06 y B06 la distancia es de

025, en el lote A06 y D06 la distancia es de 0,43 de un grupo, en el lote A05 y B05 la distancia es de 0,55, en el lote A05 y C05 es de 0,68 mientras que la distancia en los dos grupos en los lotes A05 y A06 la distancia es de 1,82.

4.7. COMPONENTES PRINCIPALES

Con el análisis de componentes principales podremos observar las relaciones entre las variables de Preaviso, Severidad, Clima clasificadas éstas por años y como influyen unas en el resultado de las otras.

Como puede observar la primera componente principal (CP1) (véase figura 4.89) separa las variables de Preaviso, Severidad siendo éstas III, Floración, Estado Evolutivo y Cosecha de las variables de Clima, por tanto la mayor variabilidad se dan con éstas variables. Es decir en éstos lotes la enfermedad de Sigatoka se encuentra más asociada. En el año 2005 decimos que se encontraban con un mayor nivel de concentración de la enfermedad de la Sigatoka en su hoja III, estado evolutivo, en la etapa de floración y cosecha.

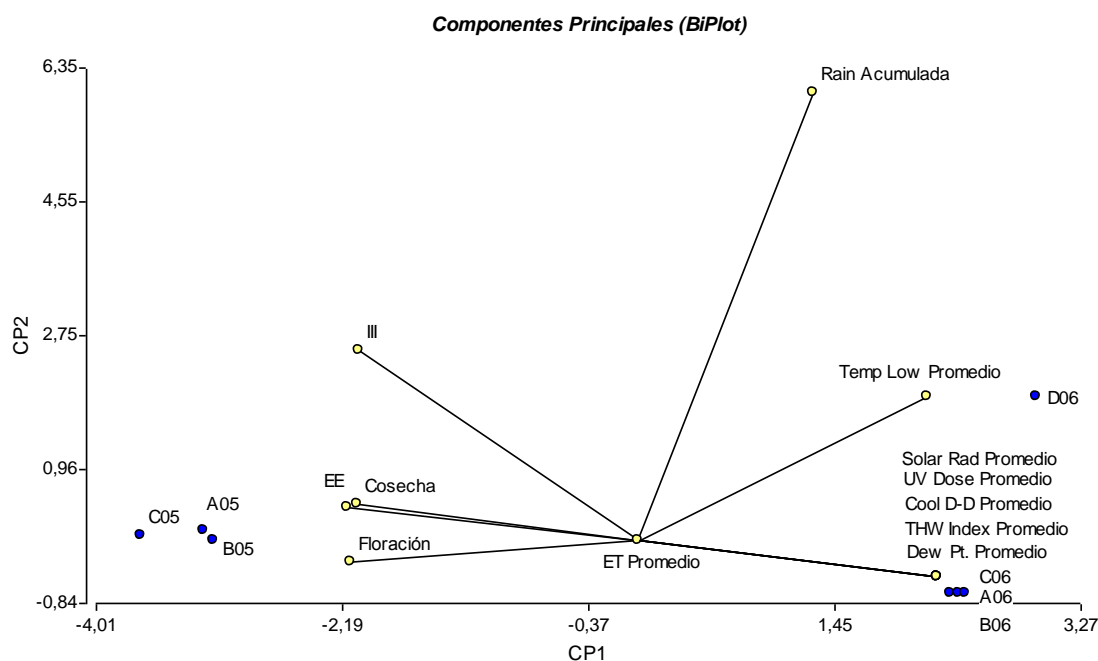


Figura 4.89 BIPLLOT de Componentes Principales

Los lotes A06, B06, C06 se encuentran más asociadas a las variables de clima como es el caso de Solar Rad, UV Dose, Cool D-D, THW Index, Dew Pt, mientras que el lote D06 se encuentra más asociada a la variable de Temperatura como es el caso de Temp Low. En el año 2006 decimos que se encuentra más relacionada con el clima.

En los lotes A05, B05 y C05 las variables Solar Rad, Temp Low, UV Dose, Cool D-D, THW Index, Dew Pt. se encuentran inversamente relacionadas, es decir en el año 2005 debido a la concentración baja de las variables de clima afectó desfavorablemente, aumentando la concentración de la enfermedad en su Hoja III, Estado Evolutivo, Floración y Cosecha.

En los lotes A06, B06, C06 y D06 las variables III, Estado Evolutivo, Cosecha y Floración se encuentran inversamente relacionadas, es decir en el año 2006 existió una concentración baja de la enfermedad de la Sigatoka y una mayor concentración de las variables de clima, decimos entonces que la Radiación Solar, Rayos Ultravioletas, los días frescos, la temperatura, humedad y viento, el punto de condensación afectaron favorablemente a éstos lotes para que exista una menor concentración de la enfermedad en su estado de Preaviso y Severidad.

Auto vectores		
Variables	e1	e2
III	-0,3	0,37
EE	-0,31	0,06
Floración	-0,31	-0,04
Cosecha	-0,3	0,07
Rain Acumulada	0,19	0,87
Temp Low Promedio	0,31	0,28
THW Index Promedio	0,32	-0,07
Cool D-D Promedio	0,32	-0,07
Solar Rad Promedio	0,32	-0,07
UV Dose Promedio	0,32	-0,07
ET Promedio	0	0
Dew Pt. Promedio	0,32	-0,07

Tabla II Auto vectores

La variable de clima ET Promedio o Evotranspiración no tiene incidencia alguna en los lotes respecto a la enfermedad de la Sigatoka ya que tiene valores en los vectores e1 y e2 de ceros (vea la Tabla II). La variable Rain Acumulada es muy importante en la creación de la componente principal dos

(CP₂) teniendo un valor de 0,87 en el vector e₂, no afectando directamente a las variables de preaviso y severidad sino moderando las variables de clima como es el caso de Humedad, Temperatura, Días Frescos.

Los resultados de la Tabla III, señalan que con las dos primeras componentes es posible explicar el 98% de la variación total. Los Autovectores descritos en la Tabla II muestran los coeficientes con que cada variable original fue ponderada para conformar las CP₁ y CP₂. Aquí podemos visualizar que al construir la CP₁ las variables III, EE, Floración y Cosecha reciben los pesos negativos más altos y las variables THW Index, Cool D-D, Solar Rad, UV Dose, Dew Pt. y Tem Low reciben el peso positivo más alto.

Autovalores			
Lambda	Valor	Proporción	Prop. Acum.
1	9,85	0,9	90,00%
2	0,87	0,08	98,00%
3	0,15	0,01	99,00%
4	0,11	0,01	100,00%
5	0,01	0	100,00%
6	0	0	100,00%
7	0	0	100,00%
8	0	0	100,00%
9	0	0	100,00%
10	0	0	100,00%
11	0	0	100,00%
12	0	0	100,00%

Tabla III Autovalores

La Componente Principal Tres (CP₃) explica el 1%, mientras que la mayor explicación se encuentra en las dos primeras componentes explicando así en la componente principal uno (CP₁) el 90%, mientras que con la componente principal dos (CP₂) el 8%, por lo que no es necesario realizar la componente principal tres.

4.8. ANÁLISIS DE CORRESPONDENCIA

Para realizar el análisis de correspondencia procedemos a categorizar los datos, debido a la cantidad de observaciones de las variables estudiadas se procedió a obtener el vector de medias, utilizamos también los cuartiles (vea Tabla IV) para la categorización.

Categorización			
Variable	Q1	Q2	Q3
III	560,5	655	942,5
EE	955,23	1186,75	1748,5
Floración	4,98	6,04	6,71
Cosecha	15,43	20,2	23,03
Rain Acumulada	2,03	25,18	85,28
Temp Low Promedio	23,55	25,13	26,94
THW Index Promedio	25,29	27,67	30,62
Cool D-D Promedio	0,11	0,15	0,18
Solar Rad Promedio	116,15	132,9	160
UV Dose Promedio	0,2	0,24	0,36
ET Promedio	0,04	0,05	0,06
Dew Pt. Promedio	21,16	22,46	23,41

Tabla IV Cuartiles para Categorización

La categorización se la realizó de la siguiente manera: Muy Baja (MB) para la observación promedio que sea menor o igual al cuartil uno $\bar{x}_i \leq Q_1$, Baja (B) para la observación promedio que sea menor o igual al cuartil dos y mayores que el cuartil uno $Q_1 < \bar{x}_i \leq Q_2$, Alta (A) para la observación promedio que sea menor o igual al cuartil tres y mayor que el cuartil dos $Q_2 < \bar{x}_i \leq Q_3$ y Muy Alta (MA) para la observación promedio que sea mayor que el cuartil tres $\bar{x}_i > Q_3$.

En anexo VI podemos observar la matriz con las categorizaciones realizadas a la observación de las variables. Para la variable III del lote A05 tenemos un promedio de 921.56, ésta observación se encuentra entre $Q_2 < 921.56 \leq Q_3 = 550.50 < 921.56 \leq 942.50$ su categorización es Alta (A III) donde III es el nombre de la variable; para la variable Dew Pt. en el lote D06 tenemos un promedio de 20.59, ésta observación es mayor que el cuartil tres, $20.59 > Q_3$ en éste caso $20.59 > 23.41$ por lo tanto se le asignaría una categoría de Muy Alta (MA Dew Pt.) donde Dew Pt. es el nombre de la variable.

En el figura 4.9 de Análisis de Correspondencia surgiere en su primer eje 1 con una inercia del 52,84% (vea tabla V), observamos en el lote C05 que existió una relación Alta de infección de la enfermedad de Sigatoka en la variable Cosecha, una Muy Alta infección en la variable de Floración debido a

la relación con el clima con Muy Baja Precipitación, Muy Baja Temperatura, Muy Baja Radiación Solar, Baja Condensación, Muy Baja Evotranspiración, Muy Bajos Días Frescos.

	Autovalor	Inercias	Ji cuadrado	(%)	% acumulado
1	0,9	0,81	485,84	52,84	52,84
2	0,45	0,2	119,9	13,04	65,88

Tabla V Ji Cuadrado

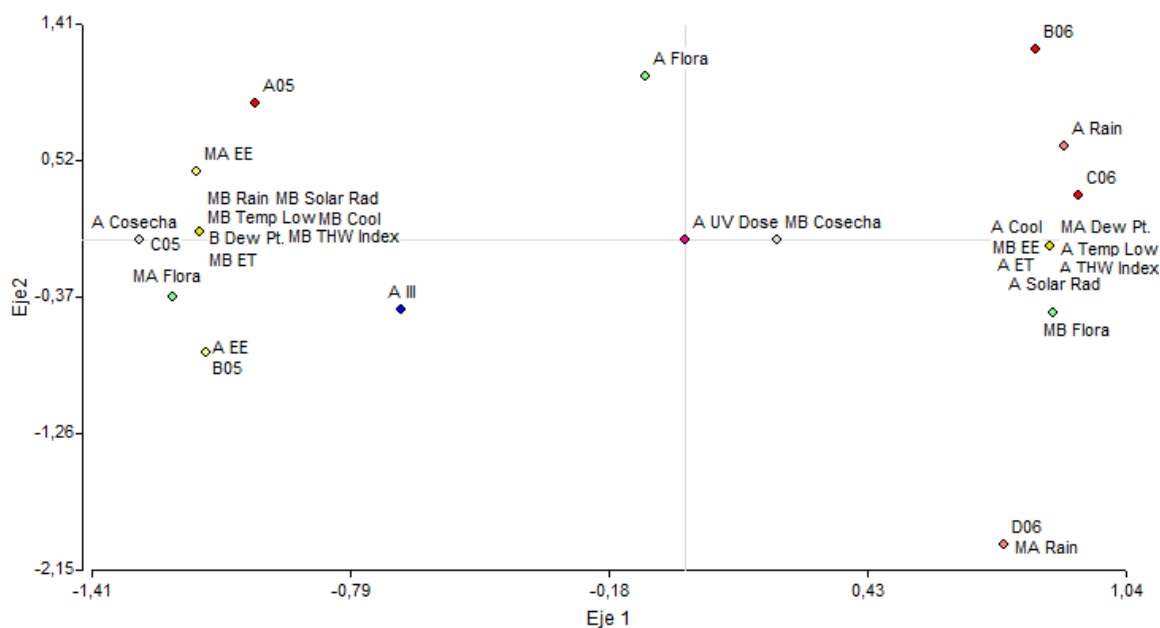
En el lote B05 existió una relación Alta de infección en el Estado Evolutivo debido a la Baja Condensación, Muy Baja Evotranspiración, Muy Baja Temperatura, Muy Baja Precipitación, Muy Baja Radiación Solar, Muy Baja Temperatura, Humedad y Viento.

En el lote D06 existió una relación Muy Alta de Precipitación, Altos días frescos, Altas Temperaturas, Alta Evotranspiración, Alta Radiación Solar, Alta relación entre Temperatura, Humedad y Viento por lo que se presento una Muy Baja Infección en la variable Estado Evolutivo, Muy Baja Infección en la variable Floración. El grado de infección en los lotes B06, C06 son parecidos al lote D06 donde existió una Muy Baja Infección en el Estado Evolutivo con Alta Precipitación, Altos días frescos, Muy Altos puntos de condensación, Altas Temperaturas Mínimas, Alta Radiación Solar, Alta Evotranspiración.

También nos indica que existe una Alta Infección en la variable de Floración por estar poco relacionada con una Alta Radiación Ultravioleta.

En la figura 4.9 sugiere con una inercia del 13,04% que en el lote A05 existió una Alta Infección en Floración al igual que el lote B06, el lote C05 está estrechamente relacionado con una Alta Infección el Cosecha junto con una Muy Baja Precipitación, Muy Baja Radiación Solar, Muy Baja Temperatura, Muy Bajos Días Frescos, Muy Baja Evotranspiración y Muy Baja relación entre Humedad, Temperatura y Viento.

Figura 4.9 Análisis de Correspondencia



En el lote B05 se encuentra estrechamente relacionado con una Alta Infección en el Estado Evolutivo y una Muy Baja Infección en la Floración.

CAPÍTULO V
CONCLUSIONES Y
RECOMENDACIONES

5.1. CONCLUSIONES

- El hongo *Mycosphaerella fijensis*, agente causante de la enfermedad de la Sigatoka negra, su mayor desarrollo se debe principalmente a las variables relacionadas con el clima, como es el caso de Precipitación ó la acumulación de Lluvia, la Temperatura, Humedad, Viento entre las principales.
- Los lotes del año 2005 de la hacienda en estudio son muy parecidos en cuanto al grado de la enfermedad de la Sigatoka negra presentando una mayor cantidad de infección en su etapa de preaviso y severidad por lo que forman un solo grupo; al igual que los lotes del año 2006 son parecidos a diferencia que estos presentan una menor cantidad de infección en su etapa de preaviso y severidad por lo que forman un solo grupo, según resultado del análisis de conglomerados.
- El grado de infección con la enfermedad de la Sigatoka negra en los años 2005 y 2006 son diferentes debido al comportamiento de las variables de clima.
- Las variables que tienen mayor relación en la etapa de Preaviso y Severidad son: III, Estado Evolutivo y Floración, Cosecha

respectivamente con las variables de clima Rain Acumulada (Precipitación), Temp Low Promedio, THW Index Promedio, Cool D-D Promedio, Solar Rad Promedio, UV Dose Promedio, ET Promedio, Dew Pt. Promedio.

- Según los resultados de las componentes principales decimos que en los lotes A05, B05 y C05 las variables Solar Rad, Temp Low, UV Dose, Cool D-D, THW Index, Dew Pt. se encuentran inversamente relacionadas, es decir en el año 2005 debido a la concentración baja de las variables de clima afectó desfavorablemente, aumentando la concentración de la enfermedad.
- Según los resultados de las componentes principales decimos que en los lotes A06, B06, C06 y D06 las variables III, Estado Evolutivo, Cosecha y Floración se encuentran inversamente relacionadas, es decir en el año 2006 existió una concentración baja de la enfermedad de la Sigatoka y una mayor concentración de las variables de clima, decimos entonces que la Radiación Solar, Rayos Ultravioletas, los días frescos, la temperatura, humedad y viento, el punto de condensación afectaron favorablemente a estos lotes para que exista una menor concentración de la enfermedad en su estado de Preaviso y Severidad.

- Fue necesario dos componentes principales para explicar el 98% de variabilidad total en las observaciones de Preaviso, Severidad y Clima.
- También se analizaron las variables por el método del análisis de correspondencia obteniendo una inercia acumulada en los dos ejes del 65.88%, con datos categorizados como es el caso de (MB) muy Baja, (B) Baja, (A) Alta, (MA) Muy Alta, encontrando así que en los lotes A05, B05 y C05 existió una Alta Infección en el Estado Evolutivo, Alta Infección en la Cosecha, Muy Alta Infección en la Floración debido a la Muy Baja Precipitación o lluvia, Muy Bajos valores de Temperaturas, Muy Bajos días Frescos, Muy Baja Radiación Solar, existió también Muy Baja relación entre la Humedad, Temperatura y Viento, Muy Baja Evotranspiración. Mientras que en los lotes A06, B06, C06 y D06 existió Muy Baja Infección en el Estado Evolutivo, Muy Baja Infección en la Floración debido a la Alta lluvia, Alta Radiación Solar, Alta Temperatura Mínima, Alta Relación entre Temperatura, Humedad y Viento, Alto Días Frescos, Alta Evotranspiración.

5.2. RECOMENDACIONES

- Realizar un estudio multivariado adicional en otras haciendas para el análisis de datos biológicos recopilados hasta la fecha sobre el desarrollo de la enfermedad de la Sigatoka negra y su relación con el clima, para así poder construir un sistema de alertas a través de un modelo de predicción.
- Llevar un registro de datos de los parámetros de clima, en este caso tener un equipo de medición automático en el cual se guarde información de precipitación, humedad, temperatura, viento, condensación, etc.
- Utilizar también una mayor cantidad de plantas para el análisis de la enfermedad de la Sigatoka negra como es en su estado de Preaviso y Severidad ya que es importante seguir realizando trabajos de investigación.
- Se recomienda realizar ensayos a través de diseños experimentales sean estos tratamientos de suelos, profundidad de sembrío, canales de riego, etc. para observar en qué o cuáles ensayos existe una menor incidencia de la enfermedad de la Sigatoka negra.

ANEXOS

ANEXO I

Correlación de Pearson					
Correlación:	III	IV	EE	Floración	Cosecha
Rain Acumulada	-0,347	-0,194	-0,317	-0,37	-0,061
	0,012	0,167	0,022	0,007	0,67
Rain Promedio	-0,338	-0,189	-0,308	-0,367	-0,054
	0,014	0,18	0,026	0,007	0,705
Temp Out Acumulada	-0,214	-0,08	-0,194	0,114	0,014
	0,127	0,572	0,167	0,42	0,92
Temp Out Promedio	-0,527	-0,437	-0,511	-0,257	-0,27
	0	0,001	0	0,065	0,053
Temp Hi Acumulada	-0,215	-0,08	-0,194	0,115	0,016
	0,126	0,571	0,168	0,418	0,909
Temp Hi Promedio	-0,531	-0,439	-0,512	-0,258	-0,265
	0	0,001	0	0,065	0,058
Temp Low Acumulada	-0,214	-0,08	-0,195	0,114	0,013
	0,129	0,572	0,167	0,421	0,929
Temp Low Promedio	-0,523	-0,435	-0,509	-0,255	-0,271
	0	0,001	0	0,068	0,052
Wind Chil Acumulada	-0,214	-0,08	-0,194	0,114	0,014
	0,127	0,572	0,167	0,42	0,92
Wind Chill Promedio	-0,527	-0,437	-0,511	-0,257	-0,27
	0	0,001	0	0,065	0,053
Heat Index Acumulada	-0,297	-0,158	-0,272	0,03	-0,044
	0,032	0,264	0,051	0,834	0,758
Heat Index Promedio	-0,548	-0,45	-0,523	-0,293	-0,287
	0	0,001	0	0,035	0,039
THW Index Acumulada	-0,297	-0,158	-0,272	0,03	-0,044
	0,032	0,264	0,051	0,834	0,758
THW Index Promedio	-0,548	-0,45	-0,523	-0,293	-0,287
	0	0,001	0	0,035	0,039
THSW Index Acumulada	-0,31	-0,169	-0,283	0,019	-0,051
	0,025	0,23	0,042	0,894	0,721
THSW Index Promedio	-0,547	-0,446	-0,52	-0,291	-0,285
	0	0,001	0	0,037	0,041
Cool D-D Acumulada	-0,455	-0,332	-0,436	-0,12	-0,151
	0,001	0,016	0,001	0,395	0,284
Cool D-D Promedio	-0,527	-0,445	-0,519	-0,238	-0,296
	0	0,001	0	0,089	0,033
Solar Rad Acumulada	-0,38	-0,222	-0,299	-0,097	-0,093
	0,005	0,113	0,032	0,496	0,514

Solar Rad Promedio	-0,504	-0,366	-0,401	-0,296	-0,221
	0	0,008	0,003	0,033	0,115
Solar Energy Acumulada	-0,38	-0,222	-0,299	-0,097	-0,093
	0,005	0,113	0,032	0,496	0,514
Solar Energy Promedio	-0,505	-0,366	-0,401	-0,297	-0,22
	0	0,008	0,003	0,033	0,117
Hi Solar Rad Acumulada	-0,381	-0,21	-0,293	-0,106	-0,087
	0,005	0,135	0,035	0,454	0,541
Hi Solar Rad Promedio	-0,498	-0,343	-0,387	-0,301	-0,206
	0	0,013	0,005	0,03	0,143
UV Index Acumulada	-0,444	-0,273	-0,339	-0,306	-0,256
	0,001	0,05	0,014	0,027	0,067
UV Index Promedio	-0,489	-0,33	-0,369	-0,413	-0,318
	0	0,017	0,007	0,002	0,022
UV Dose Acumulada	-0,444	-0,273	-0,339	-0,306	-0,256
	0,001	0,05	0,014	0,027	0,067
UV Dose Promedio	-0,488	-0,33	-0,368	-0,418	-0,321
	0	0,017	0,007	0,002	0,02
Hi UV Acumulada	-0,447	-0,272	-0,34	-0,313	-0,249
	0,001	0,051	0,014	0,024	0,074
Hi UV Promedio	-0,496	-0,333	-0,372	-0,424	-0,312
	0	0,016	0,007	0,002	0,024
ET Acumulada	-0,377	-0,216	-0,301	-0,078	-0,069
	0,006	0,123	0,03	0,581	0,629
ET Promedio	-0,398	-0,247	-0,322	-0,199	-0,104
	0,004	0,078	0,02	0,156	0,465
Out Hum Acumulada	0,108	0,199	0,104	0,281	0,136
	0,445	0,157	0,464	0,044	0,336
Out Hum Promedio	0,195	0,08	0,083	-0,109	-0,172
	0,165	0,571	0,561	0,44	0,223
Dew Pt. Acumulada	-0,213	-0,085	-0,207	0,089	-0,016
	0,129	0,55	0,141	0,529	0,912
Dew Pt. Promedio	-0,509	-0,447	-0,527	-0,311	-0,341
	0	0,001	0	0,025	0,013
Wind Speed Acumulada	-0,053	0,076	-0,065	0,193	0,198
	0,706	0,591	0,645	0,169	0,16
Wind Speed Promedio	-0,152	-0,068	-0,203	-0,059	0,123
	0,283	0,632	0,148	0,68	0,385

ANEXO II

Estadística descriptiva Parte I						
Variable	n	Media	Desviación Estándar	Varianza	Error Estándar	Coficiente Variación
III	52	732,67	249,43	61018,87	34,59	34,04
IV	52	744,02	213,81	44833,56	29,65	28,74
EE	52	1414,07	656,88	423194,54	91,09	46,45
Floración	52	6,17	1,7	2,83	0,24	27,5
Cosecha	52	19,5	4,39	18,93	0,61	22,53
Rain Acumulada	52	112,18	199,18	38911,16	27,62	177,55
Rain Promedio	52	0,09	0,16	0,02	0,02	179,15
Temp Out Acumulada	52	31180,43	5126,45	25775077,37	710,91	16,44
Temp Out Promedio	52	25,36	1,92	3,61	0,27	7,57
Temp Hi Acumulada	52	31444,21	5159,42	26107749,44	715,48	16,41
Temp Hi Promedio	52	25,57	1,92	3,63	0,27	7,52
Temp Low Acumulada	52	30932,35	5095,26	25462419,4	706,59	16,47
Temp Low Promedio	52	25,16	1,91	3,59	0,27	7,6
Wind Chill Acumulada	52	31180,43	5126,45	25775077,37	710,91	16,44
Wind Chill Promedio	52	25,36	1,92	3,61	0,27	7,57
Heat Index Acumulada	52	34243,44	6345,55	39491601,65	879,97	18,53
Heat Index Promedio	52	27,83	2,96	8,6	0,41	10,64
THW Index Acumulada	52	34243,43	6345,53	39491459,12	879,97	18,53
THW Index Promedio	52	27,83	2,96	8,6	0,41	10,64
THSW Index Acumulada	52	35644,79	6781,92	45109890,38	940,48	19,03
THSW Index Promedio	52	28,96	3,27	10,51	0,45	11,3
Cool D-D Acumulada	52	180,44	58,92	3405,29	8,17	32,65
Cool D-D Promedio	52	0,15	0,04	0	0,01	27,95
Solar Rad Acumulada	52	174927,48	45634,5	2042459485	6328,37	26,09
Solar Rad Promedio	52	141,54	27,9	763,59	3,87	19,71
Solar Energy Acumula	52	7522,86	1962,57	3777597,7	272,16	26,09
Solar Energy Promedi	52	6,09	1,2	1,41	0,17	19,7
Hi Solar Rad Acumula	52	231761,37	61411,93	3698897416	8516,3	26,5
Hi Solar Rad Promed	52	187,49	38,06	1420,66	5,28	20,3
UV Index Acumulada	52	1742,51	727,51	519099,22	100,89	41,75
UV Index Promedio	52	1,4	0,52	0,27	0,07	37,05
UV Dose Acumulada	52	373,34	155,9	23838,62	21,62	41,76
UV Dose Promedio	52	0,3	0,11	0,01	0,02	37,67
Hi UV Acumulada	52	2187,31	887,7	772855,84	123,1	40,58
Hi UV Promedio	52	1,76	0,63	0,39	0,09	35,95
ET Acumulada	52	65,15	18,1	321,25	2,51	27,78
ET Promedio	52	0,05	0,01	0	0	22,91

Out Hum Acumulada	52	102348,96	13942,33	190650366,5	1933,45	13,62
Out Hum Promedio	52	83,44	2,41	5,69	0,33	2,89
Dew Pt. Acumulada	52	27236,12	4375,22	18774435,76	606,73	16,06
Dew Pt. Promedio	52	22,17	1,7	2,84	0,24	7,67
Wind Speed Acumulada	52	634,33	116,78	13375,1	16,19	18,41
Wind Speed Promedio	52	0,52	0,05	0	0,01	9,23

ANEXO III

Estadística descriptiva Parte II							
Variable	Mínimo	Máximo	Mediana	Q1	Q3	Asimetría	Curtosis
III	365	1350	655	560,5	942,5	0,79	-0,12
IV	285	1165	698	589	882,5	0,38	-0,61
EE	185	2913,75	1186,75	955,23	1748,5	0,71	-0,2
Floración	3,96	12,88	6,04	4,98	6,71	1,74	4,52
Cosecha	9,63	27,9	20,2	15,43	23,03	-0,17	-0,78
Rain Acumulada	0	680,73	25,18	2,03	85,28	2,11	3,24
Rain Promedio	0	0,55	0,02	0	0,07	2,22	3,79
Temp Out Acumulada	19074,8	38183	32159,55	29239,8	33833,4	-0,88	0,47
Temp Out Promedio	22,2	27,92	25,33	23,71	27,13	-0,19	-1,31
Temp Hi Acumulada	19226,8	38519,8	32434,6	29484	34122,1	-0,89	0,49
Temp Hi Promedio	22,4	28,15	25,54	23,88	27,33	-0,19	-1,31
Temp Low Acumulada	18932,5	37872,1	31900,5	29004,65	33556,3	-0,88	0,45
Temp Low Promedio	22,02	27,69	25,13	23,55	26,94	-0,19	-1,31
Wind Chill Acumulada	19074,8	38183	32159,55	29239,8	33833,4	-0,88	0,47
Wind Chill Promedio	22,2	27,92	25,33	23,71	27,13	-0,19	-1,31
Heat Index Acumulada	20590,2	43507,7	34227,45	31041,75	38672,4	-0,53	-0,22
Heat Index Promedio	23,33	32,06	27,67	25,29	30,62	-0,01	-1,4
THW Index Acumulada	20590,2	43507,6	34227,45	31041,75	38672,4	-0,53	-0,22
THW Index Promedio	23,33	32,06	27,67	25,29	30,62	-0,01	-1,4
THSW Index Acumulada	21303,8	45692,8	35321,4	32215,85	40438,5	-0,46	-0,31
THSW Index Promedio	24,11	33,67	28,61	25,97	32,1	0,03	-1,42
Cool D-D Acumulada	100,19	266,93	175,12	125,97	229,45	0,08	-1,49
Cool D-D Promedio	0,08	0,2	0,15	0,11	0,18	-0,21	-1,29
Solar Rad Acumulada	91947	253186	169727,5	137172	199199	0,03	-0,72
Solar Rad Promedio	102,99	190,29	132,9	116,15	160	0,44	-1,14
Solar Energy Acumula	3954,15	10888,37	7299,25	5899,13	8566,79	0,03	-0,72
Solar Energy Promedi	4,43	8,18	5,72	5	6,88	0,43	-1,14
Hi Solar Rad Acumula	119334	336620	225634	185917	274208	0,04	-0,7
Hi Solar Rad Promed	137,66	251,97	173,64	154,03	220,25	0,46	-1,28
UV Index Acumulada	730,3	3332,3	1467,15	1193,4	2212,8	0,76	-0,27
UV Index Promedio	0,93	2,46	1,12	0,95	1,68	0,87	-0,6
UV Dose Acumulada	156,43	713,97	314,3	255,73	474,1	0,76	-0,27
UV Dose Promedio	0,2	0,53	0,24	0,2	0,36	0,87	-0,62
Hi UV Acumulada	934,3	4111,6	1855,25	1524,5	2760,2	0,74	-0,33
Hi UV Promedio	1,17	3,03	1,4	1,27	2,14	0,85	-0,73
ET Acumulada	32,63	95,35	63,47	50,61	73,74	0,12	-0,7
ET Promedio	0,04	0,07	0,05	0,04	0,06	0,44	-1,35
Out Hum Acumulada	67811	120864	106263	96608	111110	-1,18	0,88

Out Hum Promedio	79,65	87,16	83,54	82,12	86,03	-0,2	-1,02
Dew Pt. Acumulada	17221,7	33261,1	28033,7	25470,35	30078,4	-0,87	0,23
Dew Pt. Promedio	19,35	24,51	22,46	21,16	23,41	-0,31	-1,1
Wind Speed Acumulada	395,9	822,3	636,5	566,3	727,4	-0,26	-0,28
Wind Speed Promedio	0,44	0,59	0,51	0,48	0,56	0,27	-1,22

ANEXO IV

KOLMOGOROV SMIRNOV				
Variables	Media	Desviación Standard	Kolmogorov-Smirnov Z	Valor p
III	732,673	249,43	1,048	0,222
IV	744,019	213,805	0,944	0,335
EE	1414,066	656,881	1,103	0,175
Floración	6,174	1,698	1,194	0,115
Cosecha	19,501	4,393	0,694	0,722
Rain Acumulada	112,182	199,184	2,329	0
Rain Promedio	0,087	0,157	2,466	0
Temp Out Acumulada	31180,431	5126,448	1,169	0,13
Temp Out Promedio	25,358	1,919	1,253	0,087
Temp Hi Acumulada	31444,213	5159,425	1,167	0,131
Temp Hi Promedio	25,573	1,924	1,244	0,091
Temp Low Acumulada	30932,352	5095,261	1,171	0,129
Temp Low Promedio	25,156	1,912	1,273	0,078
Wind Chill Acumulada	31180,431	5126,448	1,169	0,13
Wind Chill Promedio	25,358	1,919	1,253	0,087
Heat Index Acumulada	34243,438	6345,545	0,931	0,352
Heat Index Promedio	27,827	2,961	1,177	0,125
THW Index Acumulada	34243,431	6345,534	0,931	0,352
THW Index Promedio	27,827	2,961	1,177	0,125
THSW Index Acumulada	35644,794	6781,917	1,008	0,262
THSW Index Promedio	28,956	3,273	1,118	0,164
Cool D-D Acumulada	180,445	58,924	1,265	0,081
Cool D-D Promedio	0,147	0,041	1,281	0,075
Solar Rad Acumulada	174927,481	45634,501	0,863	0,446
Solar Rad Promedio	141,539	27,903	1,218	0,103
Solar Energy Acumulada	7522,858	1962,567	0,863	0,446
Solar Energy Promedio	6,087	1,199	1,224	0,1
Hi Solar Rad Acumulada	231761,365	61411,927	1,044	0,226
Hi Solar Rad Promedio	187,493	38,059	1,404	0,039
UV Index Acumulada	1742,506	727,515	1,459	0,028
UV Index Promedio	1,404	0,52	1,818	0,003
UV Dose Acumulada	373,338	155,904	1,459	0,028
UV Dose Promedio	0,3	0,113	1,875	0,002
Hi UV Acumulada	2187,31	887,699	1,434	0,033
Hi UV Promedio	1,764	0,634	1,864	0,002
ET Acumulada	65,146	18,098	0,982	0,289
ET Promedio	0,052	0,012	1,655	0,008
Out Hum Acumulada	102348,962	13942,332	1,476	0,026

Out Hum Promedio	83,437	2,408	1,065	0,207
Dew Pt. Acumulada	27236,119	4375,221	1,29	0,072
Dew Pt. Promedio	22,172	1,7	0,863	0,446
Wind Speed Acumulada	634,331	116,779	0,668	0,763
Wind Speed Promedio	0,516	0,048	1,208	0,108

ANEXO V

VECTORES DE MEDIAS												
Lote	III	EE	Flor	Cos	Rain Acumulada	Temp Low Promedio	THW Index Promedio	Cool D-D Promedio	Solar Rad Promedio	UV Dose Promedio	ET Promedio	Dew Pt. Promedio
A05	921,56	2128,02	6,6	20,14	17,8	24,84	25,77	0,12	132,04	0,26	0,05	20,52
A06	606,71	1086,48	6,03	17,47	32,83	25,5	29,09	0,16	144,25	0,31	0,05	23,48
B05	901,11	1728,42	6,91	20,19	17,8	24,84	25,77	0,12	132,04	0,26	0,05	20,52
B06	618,83	1092,16	6,19	16,89	32,83	25,5	29,09	0,16	144,25	0,31	0,05	23,48
C05	885,22	2055,04	6,95	22,62	17,8	24,84	25,77	0,12	132,04	0,26	0,05	20,52
C06	608,83	1113,97	5,97	17,33	32,83	25,5	29,09	0,16	144,25	0,31	0,05	23,48
D06	727,83	1080,58	5,95	17,26	123,27	25,89	29,09	0,16	144,25	0,31	0,05	23,48

ANEXO VI

DATOS CATEGORIZADOS						
Lote Año	III	Cat. III	EE	Cat. EE	Floración	Cat. Floración
A05	921,56	A III	2128,02	MA EE	6,6	A Flora
A06	606,71	MB III	1086,48	MB EE	6,03	MB Flora
B05	901,11	A III	1728,42	A EE	6,91	MA Flora
B06	618,83	MB III	1092,16	MB EE	6,19	A Flora
C05	885,22	A III	2055,04	MA EE	6,95	MA Flora
C06	608,83	MB III	1113,97	MB EE	5,96	MB Flora
D06	727,83	A III	1080,58	MB EE	5,95	MB Flora
Lote Año	Cosecha	Cat. Cosecha	Rain Acumulada	Cat. Rain Acumulada	Temp Low Promedio	Cat. Temp Low Promedio
A05	20,13	MB Cosecha	17,8	MB Rain	24,84	MB Temp Low
A06	17,47	MB Cosecha	32,83	A Rain	25,5	A Temp Low
B05	20,19	MB Cosecha	17,8	MB Rain	24,84	MB Temp Low
B06	16,88	MB Cosecha	32,83	A Rain	25,5	A Temp Low
C05	22,61	A Cosecha	17,8	MB Rain	24,84	MB Temp Low
C06	17,33	MB Cosecha	32,83	A Rain	25,5	A Temp Low
D06	17,26	MB Cosecha	123,27	MA Rain	25,89	A Temp Low
Lote Año	THW Index Promedio	THW Index Promedio	Cool D-D Promedio	Cat. Cool D-D Promedio	Solar Rad Promedio	Cat. Solar Rad Promedio
A05	25,77	MB THW Index	0,12	MB Cool	132,044	MB Solar Rad
A06	29,09	A THW Index	0,16	A Cool	144,248	A Solar Rad
B05	25,77	MB THW Index	0,12	MB Cool	132,044	MB Solar Rad
B06	29,09	A THW Index	0,16	A Cool	144,248	A Solar Rad
C05	25,77	MB THW Index	0,12	MB Cool	132,044	MB Solar Rad
C06	29,09	A THW Index	0,16	A Cool	144,248	A Solar Rad
D06	29,09	A THW Index	0,16	A Cool	144,248	A Solar Rad
Lote Año	UV Dose Promedio	Cat. UV Dose Promedio	ET Promedio	Cat. ET Promedio	Dew Pt. Promedio	Cat. Dew Pt. Promedio
A05	0,263	A UV Dose	0,049	MB ET	20,52	B Dew Pt.
A06	0,309	A UV Dose	0,054	A ET	23,48	MA Dew Pt.
B05	0,263	A UV Dose	0,049	MB ET	20,52	B Dew Pt.
B06	0,309	A UV Dose	0,054	A ET	23,48	MA Dew Pt.
C05	0,263	A UV Dose	0,049	MB ET	20,52	B Dew Pt.
C06	0,309	A UV Dose	0,054	A ET	23,48	MA Dew Pt.
D06	0,309	A UV Dose	0,054	A ET	23,48	MA Dew Pt.

BIBLIOGRAFIA

- [1] Manual para el manejo Integra de Sigatoka Negra (*Mycosphaerella fijiensis* Morelet) y Sigatoka Amarilla (*Mycosphaerella musicota* Leach ex Mulder) en Banano y Plátano, Luis Pérez Vicente, Organización de la Naciones Unidad para la Agricultura y la Alimentación (FAO), Instituto de Investigaciones de Sanidad Vegetal, Ministerio de la Agricultura de la República de Cuba, Septiembre de 1996.
- [2] Diseño y Análisis de Experimentos, Douglas C. Montgomery (Arizona State University), Grupo Editorial Iberoamérica.
- [3] Estadística Aplicada a los Negocios y la Economía, 3era. Edición, Allen L. Webster (Bradley University), Irwin McGrawn – Hill Editorial.
- [4] Estadística Matemática con Aplicaciones, Willian Mendenhall, Dennis D. Wackerly y Richard L. Scheaffer, 2da. Edición, Grupo Editorial Iberoamérica.
- [5] Manual del Usuario, Software InfoStat, versión 2004.
- [6] CIBE (Centro de Investigación Biotecnológicas del Ecuador), ESPOL – Guayaquil, Ecuador, <http://www.cibe.espol.edu.ec>.
- [7] Tesis de Grado “Determinación y uso de métodos estadísticos en la obtención de un modelo bioclimático para el manejo de la Sigatoka negra presente en haciendas bananeras de la provincia del Guayas”, Ing. Manuel Chóez Salazar, 2006, ESPOL.

- [8] El Cultivo del Banano, Ministerio de Agricultura y Ganadería, Sección de Cooperativas, Agr. Argelio Núñez Álvarez, 1989.
- [9] AEBE (Asociación de Exportadores de Banano del Ecuador), 2008, <http://www.aebe.com.ec>.
- [10] Formato de Tesis de Grado, Instituto de Ciencias Matemáticas, Asociación de Estudiantes de las Carreras del Instituto de Ciencias Matemáticas, 2002.
- [11] Base de Datos de la Hacienda Carmita, CIBE (Centro de Investigación Biotecnológicas del Ecuador), ESPOL – Guayaquil, Ecuador.