

La trasmisión de la epidemia de VIH/SIDA en Puerto Rico: Modelos de regresión.

por

Brenda Carolina Torres Velásquez

Tesis sometida en cumplimiento parcial de los requerimientos para el grado de

MAESTRIA EN CIENCIAS

en

MATEMÁTICAS (ESTADÍSTICA)

UNIVERSIDAD DE PUERTO RICO

RECINTO UNIVERSITARIO DE MAYAGÜEZ

2009

Aprobado por:

Julio Quintana, PhD
Presidente, Comité Graduado

Fecha

Tokuji Saito, PhD
Miembro, Comité Graduado

Fecha

Luis Gordillo, PhD
Miembro, Comité Graduado

Fecha

Zaida Torres, Prof.
Representante de Estudios Graduados

Fecha

Julio Quintana, PhD
Director del Departamento de Ciencias Matemáticas

Fecha

ABSTRACT

This research considered statistical modeling to forecast the number of deaths by AIDS in Puerto Rico using as predictors variables the number of diagnosed cases classified by transmission modes. For this study data was obtained through the Surveillance Program of the Health Department of Puerto Rico. To develop the model we implemented a regression analysis and several variables selection techniques which helped to select the best predictors among the transmission modes of deaths by AIDS. The analysis of independence helped to determine the level of association between the more important predictors.

According to these results the main modes of transmission modes are intravenous drugs use and by sexual contact.

Finally will present the best regression model to predict the total number of AIDS deaths by transmission mode and the best regression model to predict the total number of AIDS deaths by transmission mode by gender.

Furthermore, we present the correlation analysis between the variables.

RESUMEN

En esta investigación se consideró la modelación por regresión para predecir el número de muertes por SIDA en Puerto Rico usando como variables predictoras el número de casos diagnosticados clasificados por modo de transmisión.

Los datos para este estudio se obtuvieron a través del Programa de Vigilancia de SIDA del Departamento de Salud de Puerto Rico. Para desarrollar el modelo se usó análisis de regresión y varias técnicas de selección de variables, las cuales ayudaron a escoger entre los modos de transmisión los mejores predictores de muertes por SIDA. El análisis de independencia ayudó a determinar el nivel de asociación entre los predictores más importantes.

De acuerdo a estos resultados las formas principales de contagio son el uso de drogas intravenosas y el contacto sexual.

Finalmente, se presentará el mejor modelo de regresión para predecir el total de muertes por SIDA por modo de transmisión, y los modelos para los totales de muertes por SIDA por modo de transmisión, por género. También se presenta el análisis de correlación entre las variables.

DEDICATORIA

A mi padre, Ingeniero Marcelo Torres Valdez, por su ejemplo de perseverancia y su confianza en mí.

A mi madre, Abogada Brenda Velásquez Jiménez, por su apoyo, buen ejemplo y amor durante toda mi vida.

A mi esposo, Ms.C. Diego Viteri, por ser el mejor amigo y compañero en esta aventura y mi principal motivación.

A mis hermanos: Ricardo, Marcelo, Patricio, Fernando, y Catherine, y mi madrina Elisa por sus consejos y su fe en mí.

A mis sobrinos Marcelito y Nathaly, por ser siempre mi inspiración.

AGRADECIMIENTOS

A Dios, por haber iluminado mi camino día a día y haber permitido que culmine con éxito esta etapa de mi vida.

A mis padres, quienes siempre me impulsaron a continuar estudiando y a mejorar como profesional y ser humano día a día.

A mi esposo, por su comprensión y apoyo para que culmine con éxito las metas trazadas.

A mi familia, por sus frases de aliento y su confianza en mí.

Al Dr. Julio Quintana por su guía y dirección en el desarrollo de la presente investigación.

Al Dr. Luis Gordillo por sus acertados comentarios y consejos sobre el desarrollo de la tesis y sobre mi desenvolvimiento como estudiante en la Facultad de Ciencias Matemáticas.

Al Dr. Tokuji Saito por sus sugerencias sobre el análisis estadístico de los datos.

Al Dr. Edgar Acuña por las oportunidades brindadas para trabajar en sus proyectos, y su mentoría en el transcurso de la maestría.

Al Programa de Vigilancia del SIDA, particularmente a la Dra. Sandra Miranda, Directora, por facilitarnos la información necesaria para hacer de esta investigación una realidad.

Al personal administrativo del Departamento de Ciencias Matemáticas del Recinto Universitario de Mayagüez, por el apoyo brindado durante el transcurso de la maestría.

Tabla de Contenido

1. Introducción	1
2. Revisión Bibliográfica	4
2.1. VIH y SIDA	4
2.1.1. Historia	6
2.1.2. Epidemiología del VIH. Generalidades	7
2.1.3. Epidemiología de la enfermedad	7
2.1.4. Tipos de trasmisión	8
2.1.5. Período de incubación	8
2.2. Enfoques usados para estudiar la epidemia del VIH	9
2.3. Modelos matemáticos	11
2.3.1. Modelo SI	12
2.3.2. Ejemplo: Modelo matemático para la Ciudad de New York	13
3. Metodología	14
3.1. Datos	14
3.2. Modelo de predicción	16
3.2.1. Análisis de residuos	17
3.3. Análisis de multicolinealidad	17
3.4. Análisis de autocorrelación	17
3.5. Transformaciones a variables predictoras	18
3.6. Criterios de selección de variables	18
3.6.1. Regresión escalonada	19
3.7. Análisis de independencia	19
4. Análisis de los datos y resultados	20
4.1. Datos básicos	20
4.2. Análisis de Independencia	22
4.3. Análisis de regresión escalonada	24
4.3.1. Caso total	24

4.3.2.	Género masculino	26
4.3.3.	Género femenino	27
4.4.	Análisis de multicolinealidad	27
4.4.1.	Caso total	28
4.4.2.	Género masculino	30
4.4.3.	Género femenino	32
4.5.	Análisis de autocorrelación	33
4.5.1.	Caso total	34
4.5.2.	Género masculino	34
4.5.3.	Género femenino	35
4.6.	Regresión múltiple	35
4.7.	Análisis de regresión en función del tiempo y diagnósticos	36
4.7.1.	Caso total	36
4.7.2.	Género masculino	62
4.7.3.	Género femenino	83
4.8.	Comparación de resultados de transformaciones	100
5.	Conclusiones	101

Índice de Tablas

4.1. Diagnósticos y muertes por SIDA acumulado a diciembre del 2008	20
4.2. Casos de SIDA clasificados por grupo de edad y género	21
4.3. Casos de SIDA clasificados por tipo de contagio y género	21
4.4. Tabla de frecuencias por género y tipo de contagio de SIDA	22
4.5. Probabilidades condicionales para el género masculino	23
4.6. Tabla de frecuencias por género y grupo de edad	23
4.7. Probabilidades condicionales para grupo de Edad dado género Masculino . . .	24
4.8. Regresión escalonada con $\alpha = 0,05$	25
4.9. Regresión escalonada con $\alpha = 0,05$	26
4.10. Regresión escalonada con $\alpha = 0,05$	27
4.11. Correlaciones para el caso total	28
4.12. Correlaciones para el caso género masculino	30
4.13. Correlaciones para el caso género femenino	32
4.14. Análisis autocorrelación caso total	34
4.15. Análisis autocorrelación género masculino	34
4.16. Análisis autocorrelación género masculino	35
4.17. Coeficientes de determinación para las transformaciones de variables predictoras	100

Índice de figuras

4.1. Diagrama de puntos de muertes y casos diagnosticados de SIDA	22
4.2. Muertes por SIDA en el tiempo T.	39
4.3. Diagnósticos de SIDA en el tiempo T	42
4.4. Diagnóstico de SIDA por Contacto Homosexual y Bisexual en el tiempo T	45
4.5. Diagnósticos de SIDA por Uso de Drogas Intravenosas en el tiempo T	47
4.6. Diagnósticos de SIDA por Contacto Heterosexual en el tiempo T	50
4.7. Diagnósticos de SIDA por Contacto Homosexual y Uso de Drogas Intravenosas en el tiempo T	53
4.8. Diagnósticos de SIDA por Hemofilia en el tiempo T	55
4.9. Diagnósticos de SIDA por Transfusiones en el tiempo T	58
4.10. Diagnósticos de SIDA por transmisión Vertical en el tiempo T	60
4.11. Diagnósticos de SIDA por causas bajo investigación en el tiempo T	61
4.12. Diagnósticos de SIDA por transfusiones casos pediátricos en el tiempo T	62
4.13. Diagnósticos de SIDA por hemofilia casos pediátricos en el tiempo T	62
4.14. Muertes por SIDA en el tiempo T. Género Masculino	64
4.15. Diagnósticos de SIDA en el tiempo T. Género Masculino	66
4.16. Diagnósticos de SIDA por Contacto Homosexual en el tiempo T. Género Mas- culino	68
4.17. Diagnósticos de SIDA por Uso de Drogas Intravenosas en el tiempo T. Género Masculino	70
4.18. Diagnósticos por contacto homosexual y uso de drogas intravenosas en el tiem- po T. Género masculino	72
4.19. Diagnósticos por contacto heterosexual en el tiempo T. Género masculino	74
4.20. Diagnósticos por hemofilia en el tiempo T. Género masculino	76
4.21. Diagnósticos por transfusiones en el tiempo T. Género masculino	78
4.22. Diagnósticos por transmisión Vertical en el tiempo T. Género Masculino	80
4.23. Diagnósticos de SIDA por riesgo bajo investigación en el tiempo T. Género Masculino	81
4.24. Diagnósticos por transfusiones pediátricas en el tiempo T. Género Masculino . .	82
4.25. Diagnósticos por hemofilia en casos pediátricos en el tiempo T. Género Masculino	82

4.26. Diagnósticos por RBI pediátricos en el tiempo T. Género masculino	83
4.27. Muertes por SIDA en el tiempo T. Género femenino	85
4.28. Diagnósticos de SIDA en el tiempo T. Género femenino	87
4.29. Diagnóstico por contacto heterosexual en el tiempo T. Género femenino	89
4.30. Diagnósticos por uso de drogas intravenosas en el tiempo T. Género femenino .	91
4.31. Diagnósticos por transfusiones a través del tiempo T. Género femenino	93
4.32. Diagnósticos por transmisión vertical en el tiempo T. Género femenino	95
4.33. Diagnósticos por hemofilia en el tiempo T. Género femenino	96
4.34. Diagnósticos por transfusiones pediátricas en el tiempo T. Género femenino . .	97
4.35. Diagnósticos por hemofilia pediátrica en el tiempo T. Género femenino	98
4.36. Diagnósticos por RBIPED en el tiempo T. Género femenino	99

CAPÍTULO 1

Introducción

En la actualidad, el Virus de Inmunodeficiencia Humana (VIH) y el Síndrome de Inmunodeficiencia Adquirida (SIDA) siguen siendo causas importantes de preocupación en cuanto a la salud pública a lo largo de las últimas décadas a nivel mundial. Por ejemplo, en Estados Unidos, la cifra de muertes a causa del SIDA ascendió a 14,561 en el 2007 [14]. En Puerto Rico, el SIDA fue una de las principales causas de muerte durante la década de los 90 [38] y si bien en la actualidad las cifras de contagio de VIH y SIDA, así como de muerte se han reducido, el estudio de la epidemia es de especial interés para el desarrollo de programas de salud, así como de políticas públicas. La cifra acumulada de casos de contagio de VIH desde 1984 al cierre del 2008 fue de 7,416 personas. En cuanto al SIDA, desde 1981 a diciembre del 2008 se reportaron 32,875 diagnósticos, sin que se haya podido determinar el año de infección para todos los individuos. El número de muertes por SIDA en el mismo período fue de 20,473 personas, de las cuales aproximadamente 78 % de ellas eran del género masculino y 22 % del género femenino. El año crítico para las muertes por SIDA fue el año 1994 con 1,763 muertes. [33]

Muchos estudios investigativos muestran el uso de modelos estadísticos y matemáticos para interpretar el número de casos a través del tiempo y las variables más importantes inmersas en la

epidemiología de la enfermedad. Por ejemplo, Karon [23] estableció un modelo estadístico para estimar la incidencia de VIH en los Estados Unidos en el 2002 [23]. En el año 1994, Blower y Dowlatabadi[11] desarrollaron un análisis de sensibilidad para modelos complejos de enfermedades y usaron el caso del VIH como ejemplo.

Para el caso del Puerto Rico, el Programa de Vigilancia del SIDA [33] posee historiales con pacientes portadores de VIH, SIDA e incluso el número de muertes por año. Sin embargo no se han establecido modelos estadísticos que relacionen dichas variables. Por lo expresado anteriormente, se observa la enorme importancia de establecer estudios estadísticos enfocados a medir cualitativa o cuantitativamente el impacto de la enfermedad en la isla.

A la fecha no se ha documentado un estudio analítico estadístico completo de las diferentes variables inmersas en el estudio de esta epidemia. Tampoco existen estudios epidemiológicos generales que nos permitan interpretar el progreso de enfermedad a través del tiempo para poder establecer mecanismos de control desde un punto de vista global. La presente investigación es un trabajo pionero en Puerto Rico cuyos objetivos fueron:

- Construir un modelo estadístico de acuerdo al número de diagnósticos de SIDA y tipo de transmisión.
- Determinar cuáles son las formas de transmisión más importantes en la difusión de la enfermedad.
- Describir y predecir el número de muertes por SIDA con el uso del modelo.
- Determinar en que medida están asociadas distintas variables de estudio entre sí.

El estudio se desarrolló en dos partes. En la primera se obtuvo la literatura necesaria para entender la problemática de esta enfermedad por medio del Programa de Vigilancia del SIDA.

En la segunda parte del estudio se procedió al análisis de los datos usando como programa principal de análisis Minitab versión 15, y además el programa libre R como soporte. Finalmente el desarrollo de la tesis se hizo en el procesador de palabras Latex.

La estructura del trabajo se detalla a continuación. En el Capítulo 2 mencionamos los aspectos del VIH y del SIDA, tales como su importancia, generalidades, epidemiología, historia, tipos de transmisión, períodos de incubación y los múltiples enfoques usados para estudiar esta epidemia.

En el Capítulo 3 se plantean los métodos estadísticos que se usaron para analizar los datos. Se usaron conceptos de probabilidades totales y condicionales para establecer la independencia de las variables de interés. Para establecer el modelo estadístico se usó regresión múltiple y métodos de selección escalonada hacia adelante para poder determinar las variables óptimas para el modelo; también se implementó el análisis de multicolinealidad y autocorrelación para obtener un modelo lineal óptimo .

En el Capítulo 4 se presentan todos los resultados obtenidos usando las herramientas definidas en el Capítulo 3, y finalmente en el Capítulo 5 se presentan las conclusiones del estudio.

CAPÍTULO 2

Revisión Bibliográfica

2.1. VIH y SIDA

VIH son las siglas para el Virus de Inmunodeficiencia Humana (HIV - Human Immunodeficiency Virus). Las siglas SIDA corresponden al Síndrome de Inmunodeficiencia Adquirida (Acquired Immune Deficiency Síndrome - AIDS). El VIH y SIDA no son lo mismo. VIH es un virus que daña el sistema inmunológico de una persona poniéndola en riesgo de contraer enfermedades. Por otro lado, SIDA son enfermedades que tienen como síntoma el daño al sistema inmunológico.

Con respecto al VIH, las formas de contagios se dan por contacto de fluidos corporales tales como sangre, semen, fluidos genitales, leche materna, siendo las más comunes, las transfusiones de sangre, uso de drogas intravenosas y contacto sexual, y en menor proporción, mediante contacto indirecto, tales como contacto de sangre infectada en la retina, fosas nasales, boca, o en una herida abierta. El modo de actuar del virus en el cuerpo de un individuo es debilitando y destruyendo el sistema inmunológico, en su mayoría las células de tipo CD4. Estas células son

de vital importancia para el funcionamiento correcto y normal del sistema inmunológico de un individuo. Es más, de acuerdo a Perelson [32] es un hecho que el agotamiento de las células CD4 en los individuos infectados con VIH es un predictor altamente confiable para el desarrollo de SIDA. No existe cura en la actualidad para el VIH .

Cuando la infección del VIH en un individuo se encuentra en un estado avanzado se presentan el SIDA . Esto se debe a que el VIH causa una pérdida significativa de células blancas (células CD4). Al igual que el VIH no existe cura para SIDA.

Una persona puede ser contagiada de VIH y no contraer enfermedades SIDA. Es decir, portar el virus VIH no significa que el individuo padezca de SIDA[17] De igual forma, una persona puede ser diagnosticada con enfermedades que son de la categoría SIDA, sin que por eso sea portadora del virus VIH. La única diferencia es que a estas personas que no portan el virus no se las clasifica como individuo con SIDA. Las enfermedades oportunistas que caracterizan el Síndrome del SIDA son las siguientes [28]:

- Candidiasis
- Cáncer Cervical Invasivo
- Coccidioidomycosis, Cryptococcosis, Cryptosporidiosis
- Enfermedad de Cytomegalovirus
- Encefalopatía (Relacionada al VIH)
- Herpes Simple
- Histoplasmosis
- Isosporiasis
- Sarcoma de Kaposi
- Lymphoma caracterizada por nódulos linfáticos hinchados (lymphadenopathy)

- Complejo Microbacteriano Avium (*Mycobacterium avium* complex - MAC)
- Neumonía por *Pneumocystis carinii* (*Pneumocystis carinii* pneumonia - PCP)
- Neumonía Recurrente
- Leucoencefalopatía Progresiva Multifocal (*Progressive multifocal leukoencephalopathy* - PML)
- Septicemia por *Salmonella* recurrente.
- Toxoplasmosis del Cerebro
- Tuberculosis
- Síndrome de desgaste.

2.1.1. Historia

El VIH fue descubierto de manera oficial en la década de los 80. En julio de 1981 el New York Times reporta una rara forma de cáncer descubierta entre hombres homosexuales en la ciudad de New York. A esta forma de cáncer luego se la identificó como Sarcoma de Kaposi y en la actualidad es una de las enfermedades que pueden generarse por el SIDA.

Sin embargo, de acuerdo a análisis retrospectivos de sueros almacenados en todas partes del mundo, el origen de esta enfermedad [21] se ubica en Africa, donde se identifica el primer caso de infección en Zaire en 1959. En Estados Unidos, el primer caso de infección que se ha podido identificar de acuerdo a análisis retrospectivos, datos de 1968. Antes de que el VIH fuera descubierto, SIDA era una enfermedad considerada como rara ya que muy pocos casos eran reportados, presentándose en un grupo especial de individuos los que estaban o habían estado en algún proceso de inmuno supresión, por ejemplo, la quimioterapia como tratamiento de cáncer o incluso algunos tipos de cáncer en particular[17]

2.1.2. Epidemiología del VIH. Generalidades

El sistema inmunológico de un individuo se encuentra en su mayoría compuesto por células blancas, las cuales son producidas por la médula ósea del individuo. Cuando un organismo infeccioso logra ingresar al cuerpo de un individuo, tal como un virus o una bacteria, las células blancas son las que están programadas para destruir el organismo invasor. Sin embargo, existen ciertos organismos invasores que afectan directamente el sistema inmunológico, es decir, las células blancas. Una vez que el VIH ingresa a un individuo, el mismo se adhiere a la superficie de cierto tipo de células blancas, llamadas células T, desde donde desarrollan la capacidad de reproducirse. A medida que la reproducción del virus va infectando las células T, el sistema inmunológico se empieza a debilitar hasta que eventualmente se destruye por completo. De esta forma, el individuo infectado será más susceptible a contraer enfermedades y tendrá menor capacidad para combatir las. El SIDA es diagnosticado mediante un bajo recuento de células T y con la presencia de anticuerpos de VIH en la muestra de sangre del individuo.

2.1.3. Epidemiología de la enfermedad

El VIH/SIDA son 2 procesos infecciosos de una enfermedad de etiología viral que afecta el sistema inmunológico. El VIH ataca principalmente a los leucocitos (Glóbulos blancos) conocidos técnicamente como células (CD4+T). La enfermedad es clasificada como SIDA cuando el número de células CD4+T alcanza niveles iguales o menores a 200 mm^3 [32]. Dentro de los virus que atacan el sistema inmunológico, el VIH-1 y VIH-2 son lentivirus que se encuentran dentro del grupo de Retrovirus [2]. Ambas partículas pueden transmitir el SIDA, sin embargo el VIH-2 es originario de África y su porcentaje de transmisión sexual comparado con el VIH-1 es cinco veces menor, por lo que posee menor trascendencia fuera de África por su poca capacidad de difusión [39] [8] Los estudios epidemiológicos del VIH-1 se han enfocado en parámetros

biológicos como por ejemplo: tipos de trasmisión, período de incubación, período infeccioso, y supervivencia de pacientes de SIDA [3].

2.1.4. Tipos de trasmisión

Los tipos de trasmisión del VIH se dividen en 4 grupos: contacto sexual, parenteral, perinatal o trasmisión vertical y cutáneo [6] y [8]. Dentro de los métodos de trasmisión sexual el virus puede ser diseminado por parejas homosexuales y heterosexuales. El contacto parenteral se refiere al uso de drogas intravenosas, transfusiones de sangre, picaduras con agujas infectadas. La trasmisión vertical se refiere a la trasmisión de la madre al hijo y la cutánea por medio de membranas mucosas y la piel. En general todos los fluidos corporales que incluyen sangre, semen, secreciones vaginales, leche materna, plasma y saliva han sido documentados como fuentes de trasmisión [8]. Dependiendo del país o región en estudio así como del comportamiento de la población los porcentajes del tipo de trasmisión pueden variar. Por ejemplo, referencias estadísticas de trasmisión en Nueva York y Puerto Rico, muestran que el mayor porcentaje de trasmisión de VIH se debe al uso de drogas intravenosas [9] y [33].

2.1.5. Período de incubación

El período de incubación se define como el tiempo en que el individuo adquiere el virus (VIH-1) hasta la manifestación de los síntomas de SIDA. El tiempo puede variar dependiendo de los individuos y de la tasa de progreso de la enfermedad. Por ejemplo se ha observado que la tasa de progresión de la enfermedad en hombres homosexuales es baja en los primeros 12 a 36 meses para luego incrementarse. En general la media de incubación de una persona sexualmente activa varía entre los 7 a 10 años [3]. Un factor determinante en la relación entre la incidencia de VIH y SIDA es la probabilidad de la distribución del período de incubación. Estudios en transmi-

siones perinatales muestran un menor período de incubación comparado con infantes que han adquirido el VIH vía transfusiones de sangre. Sin embargo Lawrence et al.,[25], menciona que en una transmisión perinatal desarrollará más rápidamente SIDA, con períodos de incubación más largos. Esto puede estar relacionado a que la antigenemia persistente en infantes parece estar relacionada con el desarrollo de la enfermedad [3]

2.2. Enfoques usados para estudiar la epidemia del VIH

Para el estudio de esta importante epidemia se han descrito múltiples modelos tanto matemáticos como estadísticos a través del tiempo. A pesar de que aún es desconocida la cura para el VIH y por consiguiente, del SIDA, se han estudiado modelos que incluyen el término de complacencia para la población adulta.

Se define complacencia (complacency) como un revertimiento en el comportamiento sexual de alto riesgo, una vez que la presencia habitual de VIH se reduce a niveles tan bajos que el número de casos de SIDA se convierte en excepciones en una comunidad [6]. Este estudio fue realizado para la población de Uganda, en donde el término de complacencia fue usado en el contexto de comunidades que presentaron importantes decrecimientos en la presencia de casos con VIH.

Uno de los supuestos principales en ese estudio fue que el cambio en el comportamiento sexual depende del número de pacientes diagnosticados con SIDA. Esta dependencia ha sido estudiada ampliamente por varios autores [26].

Otro tipo de modelo matemático planteado incluye ecuaciones que consideran tanto el decrecimiento de las células CD4+T como el incremento en la carga viral observado en el transcurso

de la infección del paciente. Estas variables permiten predecir con mayor precisión la dinámica de la enfermedad en el individuo infectado. En este documento se describe la importancia de diseñar un modelo matemático para SIDA que permita entender completamente la dinámica tanto del sistema que sirve como huésped como del sistema viral. [18].

En 1988, Bailey y Estreicher [4] presentan una adaptación del modelo de dispersión espacial para influenza estudiado por Scherrer en 1985. Este estudio parte de la suposición que en cualquier tiempo t , existen x individuos susceptibles quienes son VIH negativos y y individuos infectados que son VIH positivos. La ecuación inicial planteada en esa investigación, donde n es una constante fija, es la siguiente:

$$x + y = n$$

Además, como desconocían el periodo de incubación, Bailey y Estreicher [4] ignoraron el efecto de remover los individuos catalogados como VIH positivos para pasarlos a la categoría de mortalidad por SIDA.

En la investigación mencionada Bailey y Estreicher [4] plantean un sistema de ecuaciones diferenciales, con estimación de parámetros usando datos recopilados entre 1978 y 1980 por San Francisco City Clinic, de una cohorte de 6875 homosexuales y bisexuales de género masculino que vivían en San Francisco y que fueron atendidos por enfermedades venéreas. Este sistema cubre tanto la morbilidad de VIH como la de SIDA.

Entre los resultados importantes plantean que habría un máximo de 125 nuevos casos de SIDA por mes en San Francisco para el año 1990, i.e. 1500 casos por año.

2.3. Modelos matemáticos

Debido a que las enfermedades infecciosas constituyen graves amenazas para la sociedad, la modelación matemática se ha convertido en una herramienta de vital importancia para poder determinar la dinámica de trasmisión y control de las mismas y así poder implantar programas efectivos de control e interpretar patrones epidemiológicos. En su forma más básica, existen dos tipos de modelos matemáticos: determinísticos y estocásticos[29].

Los modelos determinísticos nos permiten controlar los factores que intervienen en el estudio del proceso y generan resultados precisos, mientras que en los modelos estocásticos no es posible controlar todos los factores y por lo tanto no se pueden obtener resultados únicos ya que cada resultado posible fue obtenido usando una función de probabilidad.

Los modelos epidemiológicos parten del supuesto que un individuo se encuentra en un estado determinado de entre varios posibles estados. Por este motivo, y dependiendo del estado del individuo, una población puede ser clasificada como individuo susceptible (S), individuo infectado (I) o individuo removido (R) [29]

Por otro lado para las enfermedades infecciosas existen tres modelos epidemiológicos principales para estudiar la trasmisión, los cuales son: Modelos SI (Susceptible-Infectado), SIS (Susceptible-Infectado-Susceptible) y SIR (Susceptible-Infectado-Removido), los cuales pueden ser modelados de manera determinística o estocástica, donde el supuesto principal es que la interacción entre todos los individuos es aleatoria.

Para enfermedades infecciosas de transmisión sexual (ETS), el modelo adecuado es el SIS debido a que muy pocas enfermedades de este tipo otorgan inmunidad a un individuo luego de que este haya sido infectado. Sin embargo, para el VIH el modelo útil y más usado en la actualidad, es el modelo SI , ya que un individuo es susceptible a la enfermedad (S) y puede ser infectado (I) pero no puede ser removido ya que no existe cura para la infección. Otra versión puede darse si se considera que la remoción se deba a la muerte del individuo, en cuyo caso se estaría planteando un modelo SIR

2.3.1. Modelo SI

La forma determinística de un modelo SI está dada por la siguiente ecuación:

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\lambda I(t)S(t)}{N} \\ \frac{dI}{dt} &= \frac{\lambda I(t)S(t)}{N}\end{aligned}$$

Se observa que las variables dependientes son S e I , donde S es el número de personas susceptibles, e I el número de personas infectadas. Ambas variables están en función del tiempo.

Para la forma estocástica de este modelo se supone que cada individuo infectado tiene contacto con otro y que el mismo es aleatorio, siendo la tasa de contagio λ que representa el número de contactos por unidad de tiempo.

Sea t_k el tiempo transcurrido entre la infección del individuo $(k - 1)$ y el individuo k , para $k = 1, 2, 3, \dots$. Luego, esta variable sigue una distribución exponencial con parámetro λ . De esta forma, la variable $X(t)$ que es el número de individuos susceptibles e individuos infectados en el tiempo t , va a seguir un proceso de Poisson homogéneo, donde los estados del proceso

se identifican por $X(t) = \{S(t), I(t)\}$. De esta forma, la probabilidad de transición es la siguiente:

$$P\left(X(t + \delta) = S(t) - 1, \frac{I(t)+1}{X(t)} = S(t)\right) = \frac{\lambda I(t)}{S(t)N}$$

Este modelo es homogéneo ya que parte de la suposición de que cada individuo tiene el mismo número esperado de contactos, y por lo tanto la interacción es aleatoria.

La solución a un modelo SI traza una trayectoria curvilínea que forma la letra S , ya que al inicio del proceso el número de individuos infectados que pueden transmitir la enfermedad es bajo al igual que el número de individuos susceptibles en las últimas etapas del proceso. Por lo tanto el número de individuos infectados experimenta el mayor crecimiento durante la etapa intermedia del proceso.

2.3.2. Ejemplo: Modelo matemático para la Ciudad de New York

Para poder determinar un modelo matemático que sirva para explicar la epidemia en Puerto Rico se hizo referencia a varias publicaciones anteriores sobre el tema.

Sin embargo, dado que los datos preliminares obtenidas por el Programa de Vigilancia del SIDA [33] mostraron que de los tipos de trasmisión, el más alto es el de contagio por uso de drogas intravenosas con 38 % acumulado a diciembre del 2008, seguido por contacto heterosexual con 32 %, se estableció que el estudio efectuado en 1991 titulado "Drugs, sex and VIH: A Mathematical Model for New York City" por Blower et al [9], contiene un escenario adaptable para el caso de Puerto Rico.

CAPÍTULO 3

Metodología

En este capítulo se describe de manera detallada la forma en que se desarrolló el presente estudio. Se mencionarán las características de los datos obtenidos, los métodos para obtener el modelo de predicción, las técnicas usadas para determinar cuáles son las variables de interés y su comportamiento, así como todos los procedimientos y análisis.

3.1. Datos

Debido a la extrema importancia de esta enfermedad letal y aún sin cura, el Gobierno de Puerto Rico implementó dentro del Departamento de Salud, el Programa de Vigilancia de SIDA, el cual entre sus principales funciones tiene como objetivo recopilar las estadísticas de casos de VIH y SIDA así como de muertes debido a SIDA. Estos datos se ingresaron a un sistema de cómputo llamado eHARS.

En los reportes predeterminados del sistema de cómputo del Departamento de Estadísticas y Cómputos del Programa de Vigilancia del SIDA, se obtuvieron datos clasificados por año, y por cada una de las siguientes características [33]

- Estado de diagnóstico, por categoría de edad.
- Tipo de transmisión en adultos, por género
- Tipo de transmisión en menores de 13 años, por género
- Número de casos y de muertes por estatus de diagnóstico y categoría de edad.
- Número de casos diagnosticados por grupo de edad por género.
- Etnicidad por categoría de edad.
- Número de casos y muertes por año.

Los subgrupos por cada clasificación son los siguientes:

- Estado de diagnóstico:
 - Infectados con VIH pero no con SIDA.
- Categoría de edad:
 - Adultos y adolescentes.
 - Menores de 13 años.
- Tipo de transmisión:
 - Contacto sexual hombre a hombre. (HSH)
 - Usuarios de drogas intravenosas. (UDI)
 - Contacto sexual hombre a hombre y usuarios de drogas intravenosas. (HSHUDI)
 - Hemofilia, desórdenes de coagulación. (HEM)
 - Contacto heterosexual (CH)
 - Receptores de sangre, componentes o tejido. (TRANS)
 - Otros factores de riesgo (RBI)
 - Exposición perinatal, Casos de VIH en menores de 13 años. (TV)
 - Receptores de sangre, componentes o tejido en casos pediátricos (TPED).
 - Hemofilia, desórdenes de coagulación en casos pediátricos (HPED)

- Otros factores de riesgo en casos pediátricos (RBIPED)
- Grupo de edad:
 - Menores a 13 años (G_1).
 - Entre 13 y 14 años, inclusive. (G_2)
 - Entre 15 y 24 años, inclusive. (G_3)
 - Entre 25 y 34 años, inclusive. (G_4)
 - Entre 35 y 44 años, inclusive. (G_5)
 - Entre 45 y 54 años, inclusive. (G_6)
 - Entre 55 y 64 años, inclusive. (G_7)
 - Mayores de 65 años. (G_8)

La información detallada por año fue obtenida gracias a una solicitud especial emitida por el Departamento de Ciencias Matemáticas de la UniverSIDAd de Puerto Rico, Recinto de Mayagüez, para poder así desarrollar el presente estudio, ya que esta información no se encuentra disponible para el público en general.

3.2. Modelo de predicción

El objetivo de la investigación fue desarrollar un modelo de predicción para determinar las muertes en tiempo T . Para esto se usó un modelo de regresión múltiple para poder predecir el número de muertes por SIDA, usando como predictoras los tipos de contagio de SIDA en Puerto Rico y la variable tiempo.

3.2.1. Análisis de residuos

Para poder realizar inferencias sobre los modelos de regresión, se efectuó el respectivo análisis de residuos usando las pruebas de normalidad de Kolmogorov Smirnov y el diagrama de puntos de valores ajustados vs residuos para determinar si la varianza es constante.

3.3. Análisis de multicolinealidad

El análisis de residuos también mostró que los modelos planteados inicialmente tenían problemas de multicolinealidad. Para atender a este problema se calculó la tabla de correlaciones para observar qué variables predictoras estaban causando este problema y eliminarlas algunas de ellas del modelo lineal.

3.4. Análisis de autocorrelación

Para hacer el análisis de autocorrelación se hizo uso de la estadística de Durbin Watson D , cuya distribución depende del número de observaciones n y del número de variables predictoras k .

Se utilizará una tabla donde se obtiene una cota inferior y una cota superior para la estadística D en función de n y de k . La regla de decisión será la siguiente:

- Si la estadística D es mayor a la cota superior, entonces no existe autocorrelación.
- Si la estadística D es menor a la cota inferior, entonces existe autocorrelación, y,
- Si la estadística D se encuentra entre la cota superior e inferior, la prueba no es concluyente.

3.5. Transformaciones a variables predictoras

Debido a que el modelo de regresión se estableció en función del número de diagnósticos de SIDA por modo de contagio, fue necesario analizar la tendencia de cada variable predictora con respecto al tiempo.

En nuestro estudio, el primer paso para determinar el tipo de transformación a usarse para cada predictora fue hacer un diagrama de puntos de cada variable predictora en función del tiempo T , los cuales presentaron curvas de la forma de la función normal o de la función gamma. El siguiente paso fue efectuar ambas transformaciones a cada variable regresora y validar la mejor transformación mediante el coeficiente de determinación R^2 .

3.6. Criterios de selección de variables

Selección de variables o también llamado selección de un subconjunto de predictoras es un procedimiento estadístico que nos permite seleccionar variables para obtener el mejor modelo de regresión posible.

Entre la principales razones para usar este procedimiento se encuentran las siguientes:

- No todas las variables predictoras tienen igual importancia, por lo tanto es más eficiente trabajar con un modelo donde las variables importantes estén presentes y las que tienen poca importancia no aparezcan.
- Algunas variables pueden perjudicar la confiabilidad del modelo, especialmente si están correlacionadas con otras, luego se hace necesario eliminarlas ya que son redundantes.

- Computacionalmente es más fácil trabajar con un conjunto de variables predictoras pequeño.
- Más económico recolectar información para un modelo con pocas variables.
- Si se reduce el número de variables entonces el modelo se hace más parsimonioso. Se dice que un modelo es parsimonioso si consigue ajustar bien los datos pero usando la menor cantidad de variables predictoras posibles. Es más conveniente porque sus predicciones son más confiables y además es más robusto que el modelo original.

Con el objetivo de reducir el número de variables predictoras que expliquen las muertes por SIDA en Puerto Rico, se implantaron varios métodos de interés, de tal forma que el modelo sea óptimo y eficiente.

3.6.1. Regresión escalonada

Existen muchos métodos para seleccionar variables predictoras, entre los cuales, los de mayor importancia y difusión son los métodos de Regresión Escalonada o Stepwise. En nuestro estudio el método utilizado fue el de Regresión Escalonada hacia adelante (Stepwise Forward). El análisis de regresión escalonada fue implementado para determinar las variables que deben ser consideradas en el modelo lineal.

3.7. Análisis de independencia

Para el análisis de independencia de las variables de interés se usaron los conceptos de probabilidad condicional y probabilidad total.

CAPÍTULO 4

Análisis de los datos y resultados

En el presente capítulo se presentará un análisis de los datos y un resumen de los resultados más relevantes obtenidos en el estudio.

4.1. Datos básicos

Es importante conocer el comportamiento de los datos antes de efectuar un análisis de regresión, por lo cual se desarrollaron gráficas y tablas que ilustran los datos obtenidos. La tabla 4.1 nos presenta un resumen de casos diagnosticados de SIDA y de muertes por SIDA, clasificado por Género.

Tabla 4.1: Diagnósticos y muertes por SIDA acumulado a diciembre del 2008

	Muertes	Casos
Mujeres	16,033	4,852
Hombres	4,440	8,023
Total	20,473	32,875

La tabla 4.2 nos muestra los casos diagnosticados de SIDA clasificados por grupo de edad y género del individuo Infectado. Se observa que la mayor incidencia de casos se encuentra en individuos del grupo de edad G_4 (35 a 44 años de edad inclusive), seguido de individuos del grupo de edad $G - 3$ (25 a 34 años de edad inclusive). Esta tendencia se repite en casos clasificados por género del individuo.

Tabla 4.2: Casos de SIDA clasificados por grupo de edad y género

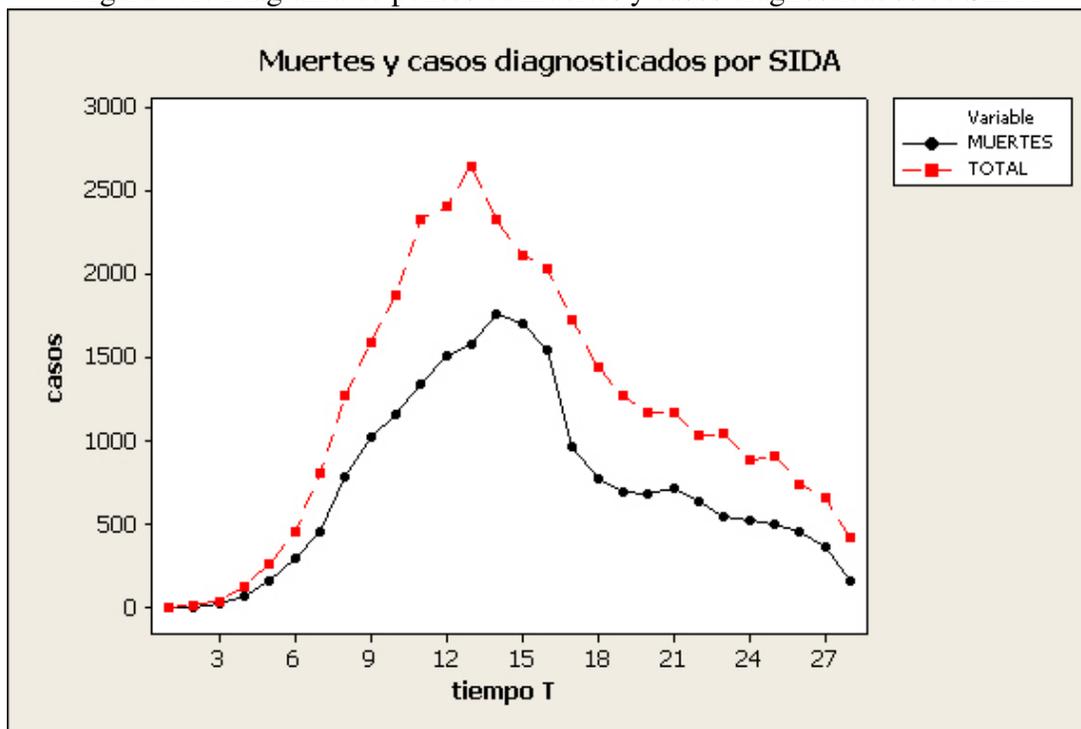
	G_1	G_2	G_3	G_4	G_5	G_6	G_7
Mujeres	229	554	2,796	2,699	1,192	413	140
Hombres	231	946	7,987	9,656	4,059	1,341	632
Total	460	1,500	10,783	12,355	5,251	1,754	772

La tabla 4.3 nos muestra las frecuencias de individuos contagiados por SIDA de acuerdo al modo en que fueron contagiados, y clasificados por género. Se observa que en el género masculino la forma de contagio principal es por uso de drogas intravenosas UDI, mientras que para el género femenino la fuente principal de infección es el contacto heterosexual.

Tabla 4.3: Casos de SIDA clasificados por tipo de contagio y género

	Hombre	Mujer	Total
HSH	5,351	0	5,351
CH	3,463	4,775	8,238
UDI	13,075	2,790	15,865
HSHUDI	2,302	0	2,302
HEM	50	6	56
TRANS	76	91	167
RBI	296	128	424
TV	207	220	427
TPED	7	4	11
HPED	9	1	10
RBIPED	16	8	24

Figura 4.1: Diagrama de puntos de muertes y casos diagnosticados de SIDA



4.2. Análisis de Independencia

En la tabla 4.4 se observan las frecuencias de casos diagnosticados por género.

Tabla 4.4: Tabla de frecuencias por género y tipo de contagio de SIDA

	HSH	CH	UDI	HUDI	HEM	TRAS	RBI	TV	TPD	HPD	RBIPD	Total
H.	5,351	3,463	13,075	2,302	50	76	296	207	7	9	16	24,852
M.	-	4,775	2,790	-	6	91	128	220	4	1	8	8,023
T.	5,351	8,238	15,865	2,302	56	167	424	427	11	10	24	32,875

Las probabilidades condicionales fueron calculadas a partir de la tabla de frecuencias anterior (tabla 4.4), siendo los resultados los que se muestran a continuación en la tabla 4.5

Tabla 4.5: Probabilidades condicionales para el género masculino

$P(HSH/Hombre)$	0.2153	$P(HSH/Mujer)$	0.000
$P(CH/Hombre)$	0.1393	$P(CH/Mujer)$	0.5952
$P(UDI/Hombre)$	0.5261	$P(UDI/Mujer)$	0.3478
$P(HSHUDI/Hombre)$	0.0926	$P(HSHUDI/Mujer)$	0.0000
$P(HEM/Hombre)$	0.0020	$P(HEM/Mujer)$	0.0007
$P(TRANS/Hombre)$	0.0031	$P(TRANS/Mujer)$	0.0113
$P(RBI/Hombre)$	0.0119	$P(RBI/Mujer)$	0.0160
$P(TV/Hombre)$	0.0083	$P(TV/Mujer)$	0.0274
$P(TPED/Hombre)$	0.0003	$P(TPED/Mujer)$	0.0005
$P(HPED/Hombre)$	0.0004	$P(HPED/Mujer)$	0.0001
$P(RBIPED/Hombre)$	0.0006	$P(RBIPED/Mujer)$	0.0010

La tabla 4.5 nos muestra que las probabilidades condicionales de cada Tipo de Contagio dado el evento de que el individuo pertenezca al género masculino, difiere en cada uno de las formas de contagio a la probabilidad condicional de cada Tipo de Contagio dado el evento de que el individuo pertenezca al complemento del género masculino, es decir, el género femenino. Estos resultados nos permiten concluir que la variable tipo de contagio y la variable género están asociadas. En la tabla 4.6 se observan las frecuencias de casos diagnosticados por género.

Tabla 4.6: Tabla de frecuencias por género y grupo de edad

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
Mujeres	229	554	2,796	2,699	1,192	413	140
Hombres	231	946	7,987	9,656	4,059	1,341	632
Total	460	1,500	10,783	12,355	5,251	1,754	772

Las probabilidades condicionales fueron calculadas a partir de la tabla de frecuencias 4.6, siendo los resultados los que se muestran en la tabla 4.7

Tabla 4.7: Probabilidades condicionales para grupo de Edad dado género Masculino

$P(G_1/Hombre)$	0.0093	$P(HSH/Mujer)$	0.0285
$P(G_2/Hombre)$	0.0381	$P(CH/Mujer)$	0.0691
$P(G_3/Hombre)$	0.3214	$P(UDI/Mujer)$	0.3485
$P(G_4/Hombre)$	0.1633	$P(HSHUDI/Mujer)$	0.1486
$P(G_5/Hombre)$	0.3885	$P(HEM/Mujer)$	0.3364
$P(G_6/Hombre)$	0.0540	$P(TRANS/Mujer)$	0.0515
$P(G_7/Hombre)$	0.0254	$P(RBI/Mujer)$	0.0174

La tabla 4.7 nos muestra que las probabilidades condicionales de cada Grupo de Edad dado el evento de que el individuo pertenezca al género masculino, difiere a la probabilidad condicional de Cada Grupo de Edad dado el evento de que el individuo pertenezca al complemento del género masculino, es decir, el género femenino. Estos resultados nos permiten concluir que la variable grupo de edad y la variable género están asociadas.

4.3. Análisis de regresión escalonada

El análisis de regresión escalonada fue utilizado para determinar cuales son las variables que deben ser consideradas en el modelo lineal.

4.3.1. Caso total

En la tabla 4.8 podemos observar las variables que mejor explican la muerte de individuos por SIDA en la población total. Para esto se uso el método de regresión escalonada con selección hacia adelante.

Tabla 4.8: Regresión escalonada con $\alpha = 0,05$

Step	1	2	3	4	5
constante		14.39	15.66	16.37	12.10
UDI	0.49	0.41	0.37	0.23	0.277
valor p	0.000	0.000	0.000	0.027	0.024
CH		2.84	1.32	0.976	0.557
valor p		0.028	0.032	0.046	0.021
T			-8.07	-12.03	-6.545
valor p			0.046	0.028	0.013
HSH				1.87	2.620
valor p				0.044	0.025
HPED					-21.022
valor p					0.043
R-Sq	94.07	94.42	94.55	95.00	95.94
R-Sq(adj)	93.99	94.35	94.40	95.03	95.05

4.3.2. Género masculino

En el método de regresión escalonada hacia adelante nuestro objetivo fue establecer las principales fuentes de infección de SIDA en el género masculino. En la tabla 4.9 se presentan los resultados obtenidos por medio del programa minitab con un $\alpha = 0,05$.

Tabla 4.9: Regresión escalonada con $\alpha = 0,05$

Step	1	2	3	4	5	6
constante	-29.72	-26.57	-15.06	-12.15	-28.53	-24.92
UDI	0.31	0.74	0.85	0.24	0.64	0.022
valor p	0.004	0.005	0.007	0.002	0.005	0.033
HSH		1.84	1.15	1.92	1.34	1.417
valor p		0.041	0.021	0.019	0.034	0.026
CH			1.31	1.64	1.13	1.529
valor p			0.026	0.036	0.044	0.037
HSHUDI				0.47	1.21	1.368
valor p				0.04	0.02	0.025
HEM					31.67	34.53
valor p					0.034	0.025
T						-3.235
valor p						0.027
R-Sq	95.1	95.81	96.03	96.15	96.6	96.73
R-Sq(adj)	94.91	95.47	95.53	95.48	95.82	95.80

Con un nivel de significancia de $\alpha = 0,05$ observamos que las principales variables para explicar las muertes por SIDA de hombres en Puerto Rico son UDI, HSH, CH, HSHUDI, HEM y T.

4.3.3. Género femenino

El análisis de regresión escalonada nos permite observar las variables de mayor importancia para predecir la variable de respuesta muerte. En la tabla 4.10 se muestran los resultados obtenidos.

Tabla 4.10: Regresión escalonada con $\alpha = 0,05$

Step	1	2	3
Constant	7.017	13.79	9.64
CH	1.061	3.407	0.959
P-Value	0.000	0.000	0.000
T		-1.08	-2.280
P-Value		0.015	0.015
TV			2.336
P-Value			0.011
R-Sq	89.18	91.67	95.24
R-Sq(adj)	88.78	90.05	94.56

Para el caso del género femenino, se observó que con un nivel de significancia de $\alpha = 0,05$ las principales variables que permiten describir las muertes de mujeres en Puerto Rico son CH, T y TV.

4.4. Análisis de multicolinealidad

Para el análisis de multicolinealidad se calculó la tabla de correlaciones entre todas las variables. Luego, se procedió a escoger las variables predictoras que tenga la menor correlación entre sí. Si dos variables tenían una correlación alta se utilizó los resultados de la regresión escalonada para poder determinar cuales fueron las variables que entraban al modelo lineal. Finalmente se calcularon los distintos modelos lineales.

4.4.1. Caso total

Tabla 4.11: Correlaciones para el caso total

TOT	MTS	TOT	HSH	CH	UDI	HEM	TRNS	RBI	TV	TPED	HPED	RPED
TOT	0,98											
HSH	0,98	0,95										
CH	0,86	0,99	0,88									
UDI	0,96	1,00	0,98	0,79								
HUDI	0,89	0,91	0,89	0,62	0,94							
HEM	0,61	0,62	0,60	0,21	0,72							
TRANS	0,47	0,49	0,46	0,05	0,60	0,91						
RBI	-0,05	0,00	0,00	0,01	-0,02	0,01	0,12					
TV	0,70	0,72	0,71	0,32	0,81	0,92	0,88	-0,01				
TPED	0,16	0,17	0,13	-0,15	0,28	0,57	0,67	0,05	0,47			
HPED	0,49	0,50	0,52	0,34	0,52	0,49	0,49	-0,07	0,45	0,12		
RPED	-0,06	-0,02	0,00	0,07	-0,05	-0,17	-0,15	0,27	-0,14	-0,10	-0,02	
T	0,13	0,18	0,17	0,48	0,05	-0,37	-0,36	0,63	-0,30	-0,33	-0,15	0,34

En la tabla 4.11 se observa el nivel de asociación lineal entre dos variables aleatorias mediante el coeficiente de correlación. Se observa por ejemplo que las correlaciones más altas se encuentran entre los siguientes pares de variables: UDI-HSH con 0.98, HSHUDI-UDI con 0.94, TV-HEM con 0.92, HSHUDI-HSH con 0.89, CH-HSH con 0.88, TV-UDI con 0.81 y UDI-CH con 0.79. Esto implica que en los modelos lineales que se analizaron, para cada pareja antes mencionada, no pueden aparecer las dos variables predictoras a la vez, sino solo una de ellas.

Para nuestros datos que las variables UDI, CH y HSH se encuentran altamente correlacionadas entre sí por parejas, por lo cual solo se puede escoger una de ellas. Adicionalmente, de acuerdo al análisis de regresión escalonada, con $\alpha = 0,025$ se observó que las principales variables para determinar el número de muertes de acuerdo a los distintos tipos de contagio de los individuos,

fueron CH, UDI y T. Al no poder ingresar las variables CH y UDI a la vez, y dado que la frecuencia de casos de contagio de SIDA en Puerto Rico por UDI es mayor a la de contagio por CH, se analizó el modelo lineal tomando sólo la variable UDI y T, y luego sólo la variable UDI, siendo los resultados los que se muestran a continuación:

Predictor	Coef	SE Coef	T	P
Constante	-49.21	69.54	-0.71	0.4860
UDI	1.25	0,07	17.82	0.0000
T	4.95	3.55	1.39	0.1760

Se observó que la variable tiempo T tuvo un valor p de 0.1760 que es mayor al nivel de significancia fijado de $\alpha = 0,05$ por lo tanto se considera que la misma no es significativa y se elimina del modelo. A continuación se presenta el análisis sin la variable predictora tiempo T.

Predictor	Coef	SE Coef	T	P
Constante	19.55	49.84	0.39	0.0398
UDI	1.26	0.07	17.61	0.000

$R - Sq = 92,3\%$	$R - Sq(adj) = 92,0\%$
-------------------	------------------------

Fuente	DF	SS	MS	F	P
Regresión	1	7,392,926	7,392,926	310.13	0.000
Error	26	619,783	23,838		
Total	27	8,012,708			

Finalmente, la ecuación de regresión para determinar las muertes por SIDA en Puerto Rico para el caso general, está dada por la siguiente ecuación de regresión:

$$MUERTES = 19.6 + 1.26 UDI$$

4.4.2. Género masculino

Tabla 4.12: Correlaciones para el caso género masculino

	Mts	Tot	HSH	UDI	HUD	CH	HEM	TRS	RBI	TV	TPD	HPD	RPD
Tot	0,98												
HSH	0,98	0,99											
UDI	0,97	0,99	0,98										
HUD	0,91	0,92	0,89	0,94									
CH	0,80	0,97	0,84	0,74	0,57								
HEM	0,66	0,67	0,62	0,73	0,82	0,16							
TRS	0,45	0,50	0,42	0,57	0,68	-0,04	0,84						
RBI	-0,06	-0,01	-0,01	-0,03	-0,12	-0,01	0,00	0,11					
TV	0,72	0,74	0,69	0,79	0,87	0,27	0,90	0,75	-0,02				
TPD	0,16	0,18	0,13	0,24	0,28	-0,13	0,32	0,53	-0,09	0,19			
HPD	0,45	0,45	0,46	0,45	0,57	0,24	0,47	0,47	-0,11	0,37	0,07		
RPD	-0,02	0,02	0,05	0,00	-0,09	0,09	-0,13	-0,05	0,26	-0,10	0,00	0,05	
T	0,09	0,13	0,17	0,04	-0,12	0,49	-0,37	-0,36	0,62	-0,30	-0,31	-0,15	0,36

En la tabla 4.12 se muestran las correlaciones entre las distintas predictoras. Es interesante señalar que las altas correlaciones se presentan entre las variables HSH-UDI con 0.98; UDI-CH con 0.74; CH-HSH con 0.84; HSHUDI-UDI con 0.94; y HEM-UDI con 0.73. Luego, de acuerdo al análisis de regresión escalonada para el género masculino se concluyó que las variables de interés para explicar las muertes en este género fueron HSH, UDI, CH, HSHUDI, HEM y T. Por otro lado, de acuerdo a las frecuencias de casos de contagio de SIDA por modo de transmisión, en este género se observó que el principal modo de contagio fue por UDI, seguido de HSH, CH y HSHUDI.

De acuerdo a los criterios mencionados anteriormente, si se selecciona la variable UDI para el modelo lineal para explicar las muertes en los hombres en Puerto Rico; por su alta correlación no puede ingresar al modelo las variables HSH y HSHUDI. La correlación de UDI con la

variable predictora CH no es considerablemente alta por lo cual puede ser considerada para el modelo, y la correlación con la variable tiempo T es de 0.04 así que se considera que no están correlacionadas. A continuación se muestran los resultados con estos modelos:

Predictor	Coef	SE Coef	T	P
Constante	32.79	47.91	0.68	0.500
UDI	0.95	0.09	10.07	0.000
CH	1.23	0.40	3.08	0.005
T	-3.79	3.12	-1.21	0.236

Se observó que el valor p para la predictora tiempo T es mayor que el nivel de significancia establecido de $\alpha = 0,05$ por lo tanto se considera que no es significativa y se elimina del modelo.

Predictor	Coef	SE Coef	T	P
Constant	-10.38	32.44	-0.32	0.052
UDI	1.01	0.08	12.72	0.000
CH	0.89	0.29	3.06	0.005

$R - Sq = 95,2\%$	$R - Sq(adj) = 94,8\%$
-------------------	------------------------

Fuente	DF	SS	MS	F	P
Regresión	2	4,767,597	2,383,798	249.21	0.000
Error	25	239,138	9,566		
Total	27	5,006,735			

Finalmente, se consideraron solo las predictoras UDI y CH ya que ambas son significativas, por lo tanto el modelo lineal es el que se menciona a continuación:

$$Muertes = -10.4 + 1.01 UDI + 0.895 CH$$

4.4.3. Género femenino

Tabla 4.13: Correlaciones para el caso género femenino

	Mts	Tot	CH	UDI	TRNS	HEM	RBI	TV	RPED	TPED	HPED
Tot	0.967										
CH	0.942	0.969									
UDI	0.924	0.959	0.864								
TRNS	0.325	0.329	0.113	0.512							
HEM	0.215	0.299	0.141	0.444	0.584						
RBI	-0.033	0.046	0.045	-0.036	0.143	-0.027					
TV	0.546	0.565	0.355	0.74	0.847	0.678	0.01				
RPED	-0.116	-0.069	-0.025	-0.143	-0.148	-0.096	0.22	-0.119			
TPED	0.027	0.037	-0.125	0.173	0.753	0.424	0.241	0.546	-0.143		
HPED	0.293	0.338	0.287	0.378	0.109	-0.101	0.101	0.274	-0.085	-0.062	
T	0.248	0.336	0.472	0.123	-0.34	-0.237	0.59	-0.296	0.178	-0.231	-0.036

En la tabla 4.13 observamos que para el caso del género femenino, las mayores correlaciones se presentan entre las variables UDI-CH con 0.86; TV-UDI con 0.74 y TV-TRANS con 0.85. De acuerdo al análisis de regresión escalonada, las principales variables para explicar las muertes de mujeres en Puerto Rico son CH, UDI, T y TV. Sin embargo, al seleccionar primero la variable predictora CH para el modelo lineal, la variable UDI no puede ingresar al modelo por no ser significativa. Por lo tanto se analizaron modelos usando las variables predictoras CH, TV y T. A continuación se presentan los resultados obtenidos:

Predictor	Coef	SE Coef	T	P
Constante	9.64	13.52	0.71	0.043
CH	0.96	0.06	15.36	0.000
TV	2.34	0.85	2.74	0.011
T	-2.28	0.88	-2.61	0.015

$R - Sq = 95,2\%$	$R - Sq(adj) = 94,6\%$
-------------------	------------------------

Fuente	DF	SS	MS	F	P
Regresión	3	355148	118383	159,35	0,000
Error	24	17830	743		
Total	27	372979			

Finalmente la ecuación que explica las muertes por SIDA en el género femenino es la que se cita a continuación:

$$MUERTES = 9.6 + 0.960 CH + 2.34 TV - 2.28 T$$

4.5. Análisis de autocorrelación

Para hacer el análisis de autocorrelación se hizo uso de la estadística de Durbin Watson, la cual se calculó para cada modelo lineal planteado. Luego se efectuó la comparación de la estadística D con una cota inferior y una cota superior obtenidas de una tabla en función de n y de k . La regla de decisión con 95 % de confianza, está dada por:

- Si la estadística D es mayor a la cota superior, entonces no existe autocorrelación.
- Si la estadística D es menor a la cota inferior, entonces existe autocorrelación, y,
- Si la estadística D se encuentra entre la cota superior e inferior, la prueba no es concluyente.

A continuación se presentan los resultados obtenidos tanto para el caso total como para los casos por género.

4.5.1. Caso total

Tabla 4.14: Análisis autocorrelación caso total

Predictoras Incluidas	Estadística Durbin Watson	Decisión
Todas	1.80	Entre cotas: no concluyente
Significativas HSH, CH, UDI, HPED, T	1.82	Mayor que cota superior: no autocorrelación
UDI, T	1.49	Entre cotas: no concluyente
UDI	1.48	Mayor que cota superior: no autocorrelación

De la tabla 4.14 se observa que el modelo que efectivamente no tiene autocorrelación es el mismo que no presenta problemas de multicolinealidad.

4.5.2. Género masculino

Tabla 4.15: Análisis autocorrelación género masculino

Predictoras Incluidas	Estadística Durbin Watson	Decisión
Todas	1.60	Entre cotas: no concluyente
Significativas HSH, UDI, HSHUDI, CH, HEM, T	1.95	Mayor que cota superior: no autocorrelación
UDI, CH, T	1.60	Mayor que cota superior: no autocorrelación
UDI, CH	1.57	Mayor que cota superior: no autocorrelación

De la tabla 4.15 se observa que el modelo que efectivamente no tiene autocorrelación es aquel que incluye las variables predictoras UDI y CH, que a la vez es el mismo que no presenta problemas de multicolinealidad.

4.5.3. Género femenino

Tabla 4.16: Análisis autocorrelación género masculino

Predictoras Incluidas	Estadística Durbin Watson	Decisión
Todas	1.50	Entre cotas: no concluyente
Significativas 1.67	no autocorrelación	Mayor que cota superior: CH, TV, T

De la tabla 4.16 se observa que el modelo que efectivamente no tiene autocorrelación es aquel que incluye las variables predictoras CH, TV y T que a la vez es el mismo que no presenta problemas de multicolinealidad.

4.6. Regresión múltiple

Al comenzar a analizar los datos desde el punto de vista de regresión múltiple, donde el número de muertes por SIDA estaba en función de todas las variables predictoras, el modelo establecido presentó los siguientes problemas:

- Multicolinealidad entre algunas variables independientes.
- Heterocedasticidad
- Autocorrelación

Estos resultados se encuentran en el apéndice A.

4.7. Análisis de regresión en función del tiempo y diagnósticos

4.7.1. Caso total

En el modelo de regresión obtenido, la muerte por SIDA se cuantifica de acuerdo al número de individuos diagnosticados de SIDA y que fueron contagiados por un modo de contagio específico. Para hallar estos valores es importante determinar el número de individuos contagiados de SIDA por los diferentes tipos de contagio establecidos. Adicionalmente, fue de interés ver la tendencia de contagio de SIDA de los individuos, por cada tipo de contagio identificado. Para analizar los dos aspectos mencionados anteriormente, se usó la técnica de transformaciones de variables predictoras. A continuación se presenta el análisis de transformación de cada variable para el caso de la población total.

Variable de Respuesta: Muertes

Predictora: Tiempo T

Modelo Original:

Al efectuar el diagrama de puntos del tiempo T vs la variable de respuesta Muertes, se observó una gráfica muy parecida a la curva normal, por lo cual el modelo para esta variable fue el siguiente:

$$Y = \alpha * e^{\beta_1 * X + \beta_2 * X^2}$$

La transformación usada para linealizar la ecuación es el siguiente:

$$Z = \ln(Y)$$

Y el nuevo modelo, ahora linealizado es el que se cita a continuación:

$$\ln(Y) = \ln(\alpha) + \beta_1 X + \beta_2 X^2$$

$$Z = \ln(\alpha) + \beta_1 X + \beta_2 X^2$$

Usando la nomenclatura, el modelo resultante es el siguiente:

$$\ln(\text{Muertes}) = \ln(\alpha) + \beta_1 T + \beta_2 T^2$$

El Contraste de hipótesis de interés se menciona a continuación:

Contraste de hipótesis:

$$H_0 : \beta_i = 0$$

vs.

$$H_1 : \exists \beta_i \neq 0$$

La ecuación de regresión resultante para la variable transformada MUERTES es la siguiente:

$$\ln(y) = 1,03 + 0,786T - 0,0235T^2$$

Predictor	Coef	SE Coef	T	P
constante	1.0264	0.4305	2.38	0.025
T	0.78564	0.06843	11.48	0.000
T ²	-0.023535	0.002290	-10.28	0.000

S = 0.705809	R-Sq = 85.3 %	R-Sq(adj) = 84.1 %
PRESS = 17.8264	R-Sq(pred) = 78.91 %	

Análisis de Varianza

Fuente	DF	SS	MS	F	P
Regresión	2	72.055	36.028	72.32	0.000
Error	25	12.454	0.498		
Total	27	84.509			
T	1	19.435			
T ²	1	52.620			

Luego se procedió a efectuar una transformación tipo gamma.

$$Z = \ln(Y)$$

Y el nuevo modelo, ahora linealizado es el que se cita a continuación:

$$\ln(Y) = \ln(\alpha) + \beta_1 * \ln(X) + \beta_2 * X$$

$$Z = \ln(\alpha) + \beta_1 \ln(X) + \beta_1 X$$

Usando la nomenclatura, el modelo resultante es el siguiente:

$$\ln(\text{Muertes}) = \ln(\alpha) + \beta_1 * \ln(\text{Year}) + \beta_2 * \text{Year}$$

El Contraste de hipótesis a probar es el siguiente:

Contraste de hipótesis:

$$H_0 : \beta_i = 0, i = 1, 2$$

vs.

$$H_1 : \exists \beta_i \neq 0$$

Usando programa Minitab, la ecuación de regresión resultante para esta variable, al igual que el análisis de varianza correspondiente, se presenta a continuación:

$$\ln(\text{Muertes}) = -0,464 - 0,338T + 4,66\ln(T)$$

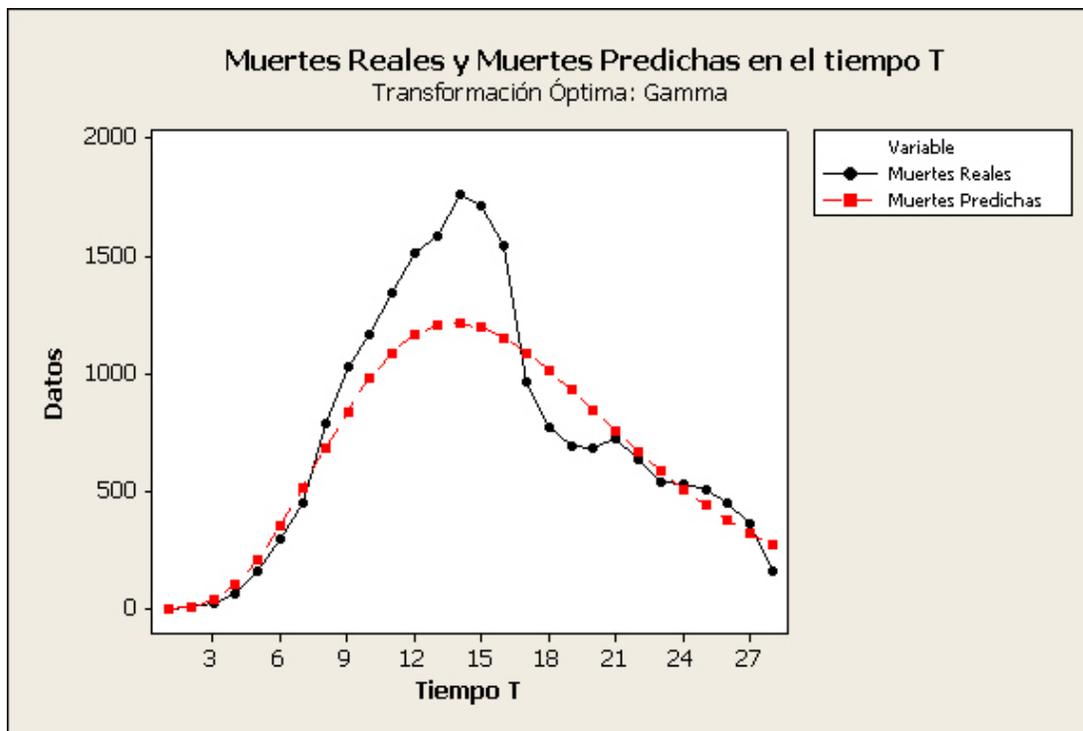
Predictor	Coef	SE Coef	T	P
constante	-0.4635	0.2378	-1.95	0.063
T	-0.33829	0.01968	-17.19	0.000
ln(T)	4.6618	0.1917	24.32	0.001

S = 0.324884	R-Sq = 96.9 %	R-Sq(adj) = 96.6 %
PRESS = 6.44691	R-Sq(pred) = 92.37 %	

Análisis de Varianza

Fuente	DF	SS	MS	F	P
Regresión	2	81.871	40.935	387.83	0.000
Error	25	2.639	0.106		
Total	27	84.509			
$\ln(T)$	1	19.435			
T	1	62.436			

Figura 4.2: Muertes por SIDA en el tiempo T.



La siguiente variable a analizar fue el Total de Casos Diagnosticados de SIDA. El primer paso es hacer la gráfica de tendencia para poder visualizar la forma de la misma y luego se implementaron las transformaciones correspondientes.

Variable de Respuesta: Total de Casos Diagnosticados de SIDA

Predictora: Tiempo T

Modelo Original:

Al efectuar el diagrama de puntos de T vs Total, observamos nuevamente una gráfica muy parecida a la curva normal, por lo tanto podemos pensar en el siguiente modelo:

$$Y = \alpha * e^{\beta_1 * X + \beta_2 * X^2}$$

La transformación usada para linealizar la ecuación es el siguiente:

$$Z = \ln(Y)$$

Y el nuevo modelo, ahora linealizado es el que se cita a continuación:

$$\ln(Y) = \ln(\alpha) + \beta_1 X + \beta_2 X^2$$

$$Z = \ln(\alpha) + \beta_1 X + \beta_2 X^2$$

Usando la nomenclatura, el modelo resultante es el siguiente:

$$\ln(TOTAL) = \ln(\alpha) + \beta_1 T + \beta_2 T^2$$

Contraste de hipótesis:

$$H_0 : \beta_i = 0, i = 1, 2$$

vs.

$$H_1 : \exists \beta_i \neq 0$$

El modelo de regresión obtenido usando el programa Minitab es el siguiente:

$$\ln(TOTAL) = 1.86 + 0.729 T - 0.0216 T^2$$

Predictor	Coef	SE Coef	T	P
constante	1.8619	0.4101	4.54	0.000
T	0.72885	0.06518	11.18	0.000
T^2	-0.021590	0.002181	-9.90	0.000

S = 0.672277	R-Sq = 84.9 %	R-Sq(adj) = 83.7 %
PRESS = 16.5616	R-Sq(pred) = 77.88 %	

Análisis de Varianza

Fuente	DF	SS	MS	F	P
Regresión	2	63.572	31.786	70.33	0.000
Error	25	11.299	0.452		
Total	27	74.871			
T	1	19.290			
T^2	1	44.282			

La siguiente transformación usada fue la de la distribución gamma, para lo cual se obtuvo la siguiente ecuación de regresión:

$$\ln(TOTAL) = 0.435 - 0.308 T + 4.33 \ln(T)$$

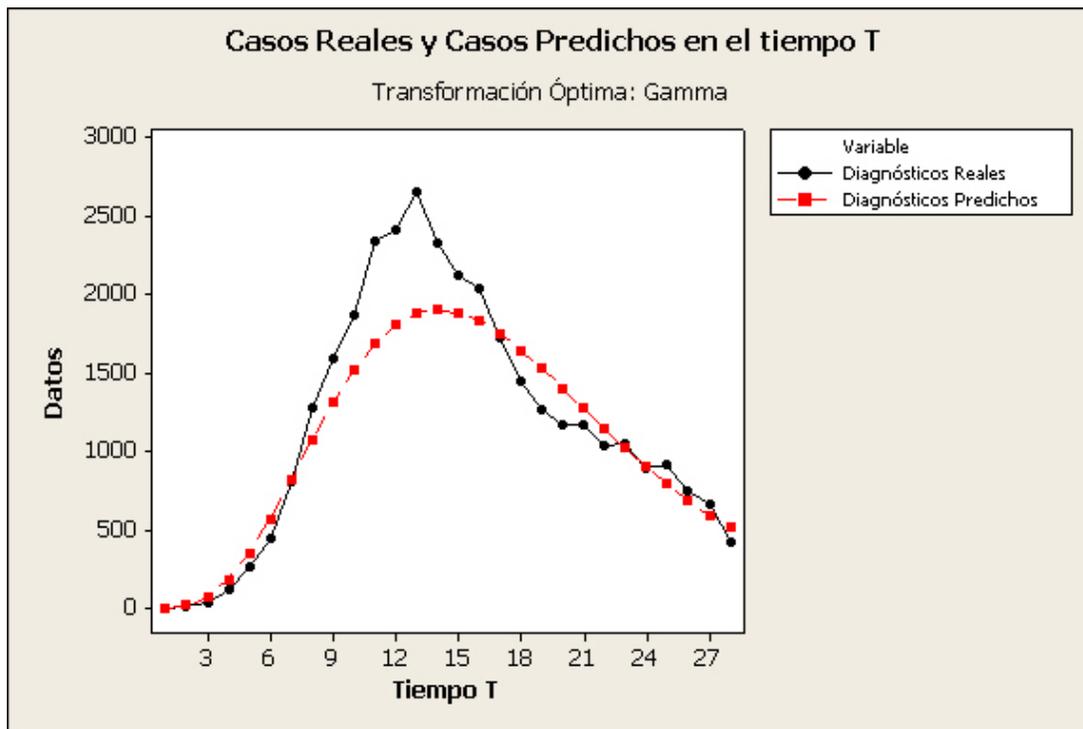
Predictor	Coef	SE Coef	T	P
Predictor	Coef	SE Coef	T	P
constante	0.4352	0.1868	2.33	0.028
T	-0.30759	0.01546	-19.90	0.000
$\ln(T)$	4.3335	0.1506	28.78	0.000

S = 0.255237	R-Sq = 97.8 %	R-Sq(adj) = 97.7 %
PRESS = 3.53823	R-Sq(pred) = 95.27 %	

Análisis de Varianza

Fuente	DF	SS	MS	F	P
Regresión	2	73.242	36.621	562.14	0.000
Error	25	1.629	0.065		
Total	27	74.871			
T	1	19.290			
T ²	1	53.953			

Figura 4.3: Diagnósticos de SIDA en el tiempo T



Variable de Respuesta: HSH

Predictora: Tiempo T

A continuación se presenta las dos transformaciones usadas para la variable predictora HSH.

Modelo Original:

Al efectuar el diagrama de puntos de T vs HSH, se observó una gráfica muy parecida a la curva normal y el modelo es el siguiente:

$$Y = \alpha * e^{\beta_1 X + \beta_2 X^2}$$

La transformación usada para linealizar la ecuación es el siguiente:

$$Z = \ln(Y)$$

Y el nuevo modelo, ahora linealizado es el que se cita a continuación:

$$\ln(Y) = \ln(\alpha) + \beta_1 X + \beta_2 X^2$$

$$Z = \ln(\alpha) + \beta_1 X + \beta_2 X^2$$

Usando la nomenclatura, el modelo resultante es el siguiente:

$$\ln(HSH) = \ln(\alpha) + \beta_1 T + \beta_2 T^2$$

Contraste de hipótesis:

$$H_o : \beta_i = 0 \quad i = 1, 2$$

vs.

$$H_1 : \exists \beta_i \neq 0$$

La ecuación de regresión obtenida por Minitab es la siguiente :

$$\ln(HSH) = 1.04 + 0.601 T - 0.0180 T^2$$

Predictor	Coef	SE Coef	T	P
constante	1.0433	0.3432	3.04	0.005
T	0.60062	0.05454	11.01	0.000
T^2	-0.017980	0.001825	-9.85	0.000

S = 0.562549	R-Sq = 84.2 %	R-Sq(adj) = 82.9 %
PRESS = 11.8311	R-Sq(pred) = 76.38 %	

Fuente	DF	SS	MS	F	P
Regresión	2	42.175	21.087	66.64	0.000
Error	25	7.912	0.316		
Total	27	50.087			
T	1	11.463			
T^2	1	30.712			

Luego, se procedió a implementar la transformación a una distribución gamma y la ecuación de regresión resultante fue la siguiente:

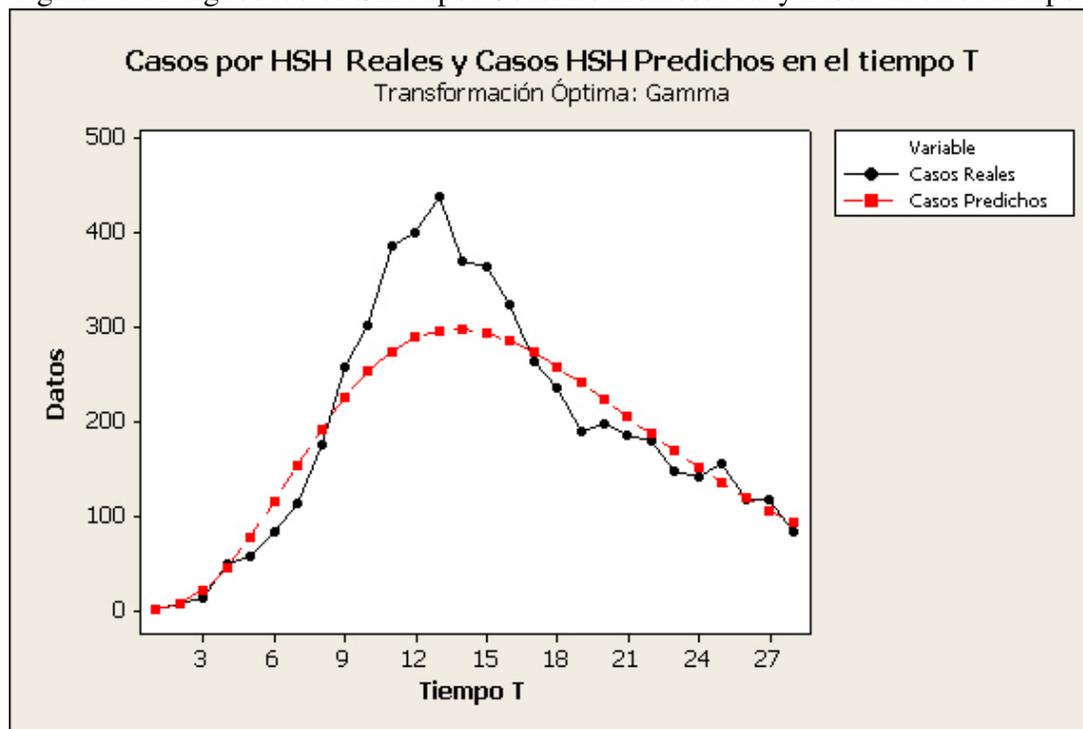
$$\ln(HSH) = -0.135 - 0.262 T + 3.60 \ln(T)$$

Predictor	Coef	SE Coef	T	P
constante	-0.1350	0.1730	-0.78	0.442
T	-0.26164	0.01432	-18.27	0.000
$\ln(T)$	3.5996	0.1395	25.81	0.000

S = 0.236420	R-Sq = 97.2 %	R-Sq(adj) = 97.0 %
PRESS = 2.43548	R-Sq(pred) = 95.14 %	

Fuente	DF	SS	MS	F	P
Regresión	2	48.689	24.345	435.55	0.000
Error	25	1.397	0.056		
Total	27	50.087			
T	1	11.463			
$\ln(T)$	1	37.226			

Figura 4.4: Diagnóstico de SIDA por Contacto Homosexual y Bisexual en el tiempo T



Variable de Respuesta: UDI

Predictora: Tiempo T

Modelo Original:

La siguiente variable transformada fue UDI a la cual se le aplicaron las dos transformaciones: normal y gamma. El modelo normal es el

$$Y = \alpha * e^{\beta_1 X + \beta_2 X^2}$$

Y la transformación que se usó fue la mencionada anteriormente:

$$Z = \ln(Y)$$

El nuevo modelo, ahora linealizado es el que se cita a continuación:

$$\ln(UDI) = \ln(\alpha) + \beta_1 T + \beta_2 T^2$$

Contraste de hipótesis:

$$H_o : \beta_i, i = 1, 2$$

vs.

$$H_1 : \exists \beta_i \neq 0$$

La ecuación de regresión así como el análisis de varianza se presenta a continuación:

$$\ln(UDI) = 1.81 + 0.658T - 0.0202 T^2$$

Predictor	Coef	SE Coef	T	P
constante	1.8051	0.4020	4.49	0.000
T	0.65786	0.06127	10.74	0.000
T ²	-0.020175	0.001990	-10.14	0.000

S = 0.559856	R-Sq = 83.0 %	R-Sq(adj) = 81.6 %
PRESS = 10.3281	R-Sq(pred) =76.68 %	

Análisis de Varianza

Fuente	DF	SS	MS	F	P
Regresión	2	36.759	18.380	58.64	0.000
Error	24	7.523	0.313		
Total	26	44.282			
T	1	4.535			
T ²	1	32.224			

La siguiente transformación usada fue a una distribución gamma, siendo la ecuación de regresión la siguiente:

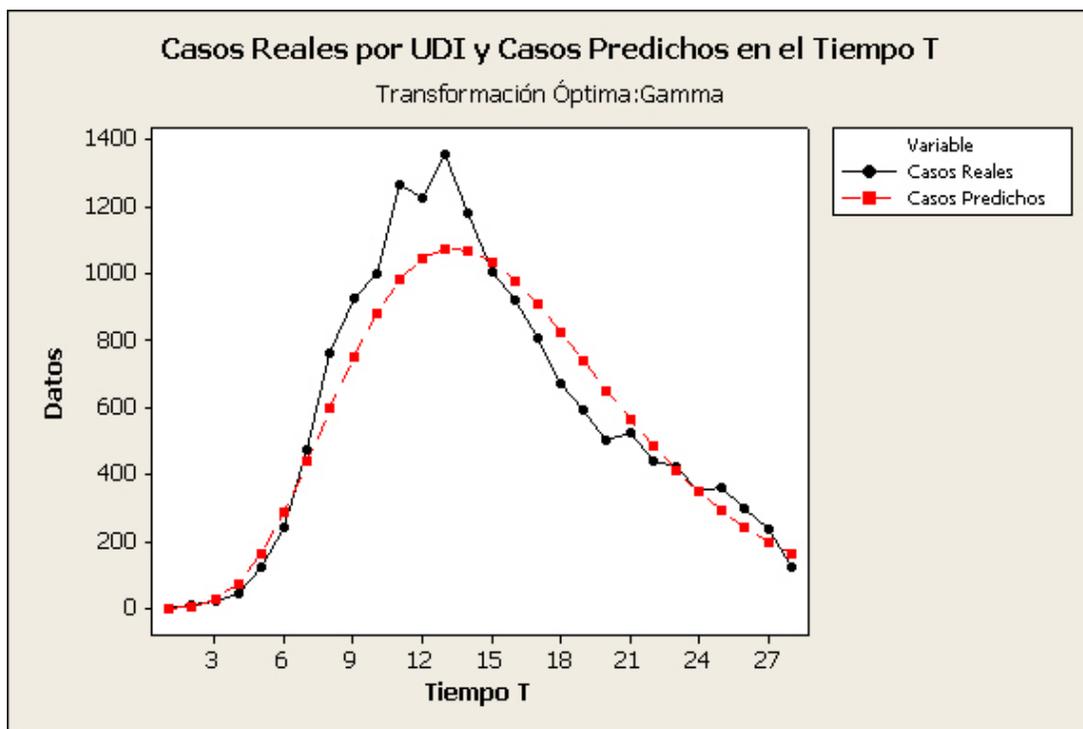
$$\ln(UDI) = - 1.48 - 0.399 T + 5.32 \ln(T)$$

Predictor	Coef	SE Coef	T	P
constante	-1.4774	0.2868	-5.15	0.000
T	-0.39871	0.01929	-20.67	0.000
$\ln(T)$	5.3164	0.2152	24.71	0.000

S = 0.250295	R-Sq = 96.6 %	R-Sq(adj)= 96.3 %
PRESS = 2.83544	R-Sq(pred)=93.60 %	

Fuente	DF	SS	MS	F	P
Regresión	2	42.778	21.389	341.42	0.000
Error	24	1.504	0.063		
Total	26	44.282			
T	1	4.535			
$\ln(T)$	1	38.243			

Figura 4.5: Diagnósticos de SIDA por Uso de Drogas Intravenosas en el tiempo T



Variable de Respuesta: CH

Predictora: Tiempo T

Modelo Original:

Al efectuar el diagrama de puntos de T vs CH, observamos una gráfica muy parecida a la curva normal, pero algo sesgada a la izquierda. Por este motivo se desarrollaron dos tipos diferentes de transformaciones para esta variable.

La primera transformación será usando la forma de una distribución gamma. El modelo se cita a continuación:

$$Y = \alpha * X^{\beta_1} * e^{\beta_2 X}$$

La transformación a usarse para linealizar la gráfica es el siguiente:

$$Z = \ln(Y)$$

Y el nuevo modelo, ahora linealizado es el que se cita a continuación:

$$\ln(Y) = \ln(\alpha) + \beta_1 * \ln(X) + \beta_2 * X$$

$$\begin{aligned} Z &= \ln(\alpha) + \beta_1 \ln(X) + \beta_2 X \\ \ln(CH) &= \ln(\alpha) + \beta_1 * \ln(Year) + \beta_2 * Year \end{aligned}$$

El Contraste de hipótesis se presenta a continuación

Contraste de hipótesis:

$$H_0 : \beta_i = 0, i = 1, 2$$

vs.

$$H_1 : \exists \beta_i \neq 0$$

La ecuación de regresión así como el análisis de varianza es el siguiente:

$$\ln(CH) = - 1.85 + 3.82 \ln(T) - 0.182 * T$$

Predictor	Coef	SE Coef	T	P
constante	-1.8510	0.5912	-3.13	0.004
$\ln(T)$	3.8245	0.4766	8.03	0.000
T	-0.18181	0.04892	-3.72	0.001

S = 0.807773	R-Sq = 86.1 %	R-Sq(adj) = 85.0 %
PRESS = 39.9088	R-Sq(pred) = 66.11 %	

Análisis de Varianza

Fuente	DF	SS	MS	F	P
Regresión	2	101.435	50.717	77.73	0.000
Error	25	16.312	0.652		
Total	27	117.747			
$\ln(T)$	1	92.424			
T	1	9.011			

La segunda opción fue usar una transformación a la distribución normal, como se desarrolló con las variables anteriores.

$$Y = \alpha * e^{\beta_1 * X + \beta_2 * X^2}$$

La transformación usada para linealizar la ecuación es el siguiente:

$$Z = \ln(Y)$$

Y el nuevo modelo, ahora linealizado es el que se cita a continuación:

$$\ln(CH) = \ln(\alpha) + \beta_1 T + \beta_2 * T^2$$

Contraste de hipótesis:

$$H_0 : \beta_i, i = 1, 2$$

vs.

$$H_1 : \exists \beta_i \neq 0$$

La ecuación de regresión es la siguiente:

$$\ln(CH) = - 1.21 + 0.856 T - 0.0233 T^2$$

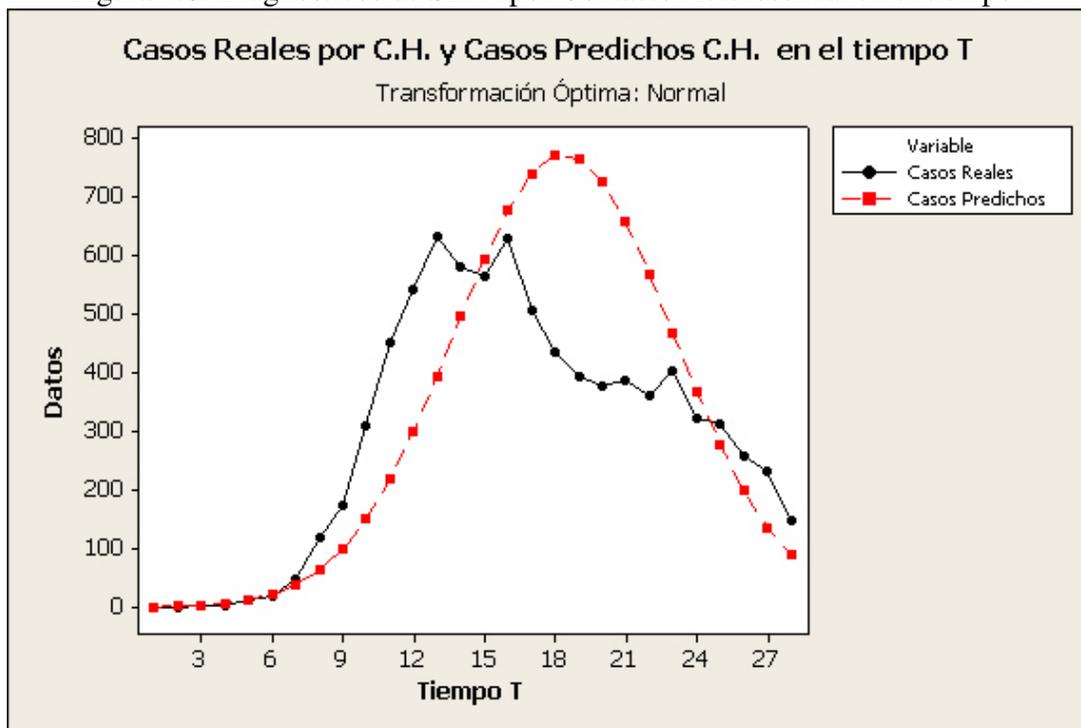
Predictor	Coef	SE Coef	T	P
constante	-1.2071	0.3175	-3.80	0.001
T	0.85593	0.05046	16.96	0.000
T^2	-0.023296	0.001689	-13.80	0.000

S = 0.520500	R-Sq = 94.2 %	R-Sq(adj) = 93.8 %
PRESS = 8.57301	R-Sq(pred) = 92.72 %	

Análisis de Varianza

Fuente	DF	SS	MS	F	P
Regresión	2	110.974	55.487	204.81	0.000
Error	25	6.773	0.271		
Total	27	117.747			
T	1	59.413			
T^2	1	51.561			

Figura 4.6: Diagnósticos de SIDA por Contacto Heterosexual en el tiempo T



Variable de Respuesta: HSHUDI

Predictora: Tiempo T

Modelo Original:

Al efectuar el diagrama de puntos de T vs HSHUDI, observamos una gráfica muy parecida a la curva normal, por lo tanto podemos pensar en el siguiente modelo:

$$Y = \alpha * e^{\beta_1 * X + \beta_2 * X^2}$$

La transformación usada para linealizar la ecuación es el siguiente:

$$Z = \ln(Y)$$

Y el nuevo modelo, ahora linealizado es el que se cita a continuación:

$$\begin{aligned} \ln(Y) &= \ln(\alpha) + \beta_1 X + \beta_2 X^2 \\ Z &= \ln(\alpha) + \beta_1 X + \beta_2 X^2 \\ \ln(HSH\&UDI) &= \ln(\alpha) + \beta_1 T + \beta_2 T^2 \end{aligned}$$

Contraste de hipótesis:

$$H_o : \beta_i = 0, i = 1, 2$$

vs.

$$H_1 : \exists \beta_i \neq 0$$

El modelo de regresión obtenido es el siguiente:

$$\ln(HSH\&UDI) = 0,794 + 0,572T - 0,0187T^2$$

Predictor	Coef	SE Coef	T	P
constante	0.7944	0.4690	1.69	0.103
T	0.57204	0.07150	8.00	0.000
T ²	-0.018680	0.002322	-8.05	0.000

S = 0.653247	R-Sq = 73.1 %	R-Sq(adj) = 70.9 %
PRESS = 15.6849	R-Sq(pred) = 58.82 %	

Análisis de Varianza

Fuente	DF	SS	MS	F	P
Regresión	2	27.847	13.924	32.63	0.000
Error	24	10.242	0.427		
Total	26	38.089			
T	1	0.223			
T^2	1	27.624			

La siguiente transformación aplicada a la variable HSHUDI fue la de la distribución gamma, siendo los resultados obtenidos los siguientes:

$$\ln(HSHUDI) = -2.59 - 0.430 T + 5.20 \ln(T)$$

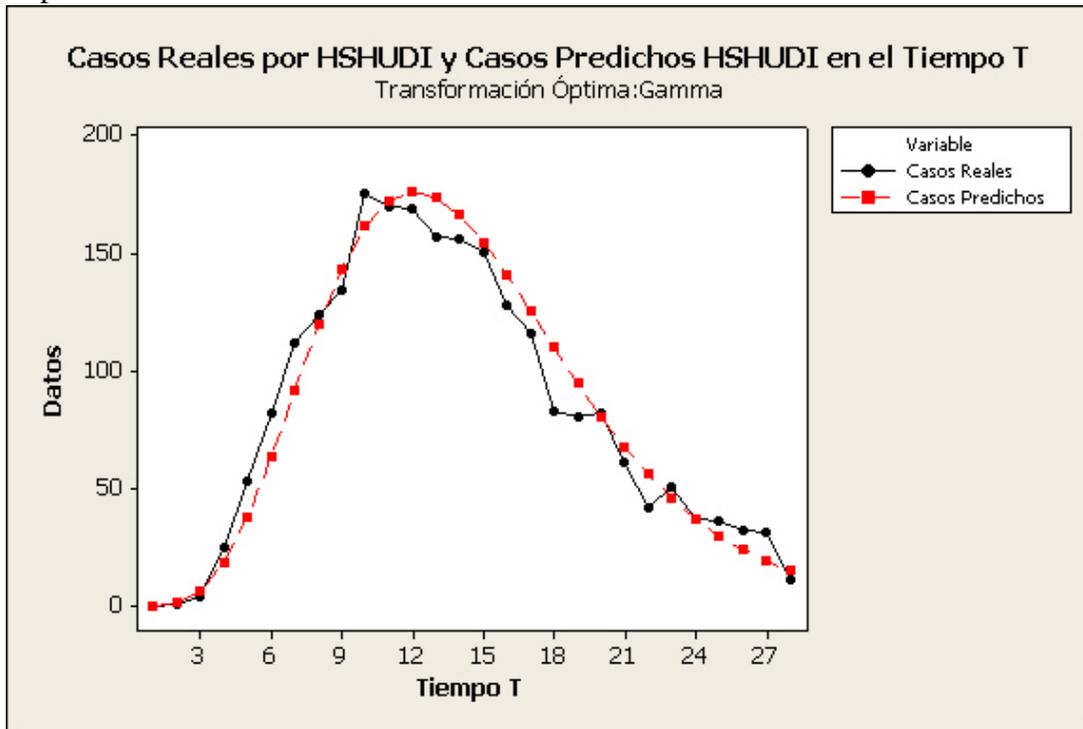
Predictor	Coef	SE Coef	T	P
constante	-2.5909	0.2627	-9.86	0.000
T	-0.42990	0.01766	-24.34	0.000
$\ln(T)$	5.2013	0.1971	26.39	0.000

S = 0.229245	R-Sq = 96.7 %	R-Sq(adj) = 96.4 %
PRESS = 1.80488	R-Sq(pred) = 95.26 %	

Análisis de Varianza

Fuente	DF	SS	MS	F	P
Regresión	2	36.827	18.414	350.38	0.000
Error	24	1.261	0.053		
Total	26	38.089			
T	1	0.223			
$\ln(T)$	1	36.605			

Figura 4.7: Diagnósticos de SIDA por Contacto Homosexual y Uso de Drogas Intravenosas en el tiempo T



Variable de Respuesta: Hem

Predictora: Tiempo T

Modelo Original:

Del diagrama de puntos entre Casos Diagnosticados por SIDA vs Tiempo T, se observa que la gráfica tiene la forma de una distribución de la familia gamma, por lo cual se implantó esa transformación. El modelo se cita a continuación:

$$Y = \alpha * X^{\beta_1} * e^{\beta_2 X}$$

La transformación a usarse para linealizar la gráfica es el siguiente:

$$Z = \ln(Y)$$

Y el nuevo modelo, ahora linealizado es el que se cita a continuación:

$$\ln(Hem) = \ln(\alpha) + \beta_1 * \ln(Year) + \beta_2 * Year$$

Contraste de hipótesis:

$$H_o : \beta_i = 0, i = 1, 2$$

vs.

$$H_1 : \exists \beta_i \neq 0$$

Los resultados obtenidos en Minitab son los siguientes:

$$\ln(HEM) = - 3.76 + 444 \ln(T) - 0.473 T$$

Predictor	Coef	SE Coef	T	P
constante	-3.7621	0.7589	-4.96	0.000
$\ln(T)$	4.4374	0.5817	7.63	0.000
T	-0.47282	0.05570	-8.49	0.000

S = 0.246884	R-Sq = 86.8 %	R-Sq(adj) = 84.6 %
PRESS = 1.36595	R-Sq(pred) = 75.36 %	

Fuente	DF	SS	MS	F	P
Regresión	2	4.8130	2.4065	39.48	0.000
Error	12	0.7314	0.0610		
Total	14	5.5444			
$\ln(T)$	1	0.4212			
T	1	4.3918			

La siguiente transformación efectuada fue a una distribución normal. Los resultados obtenidos en minitab son los siguientes:

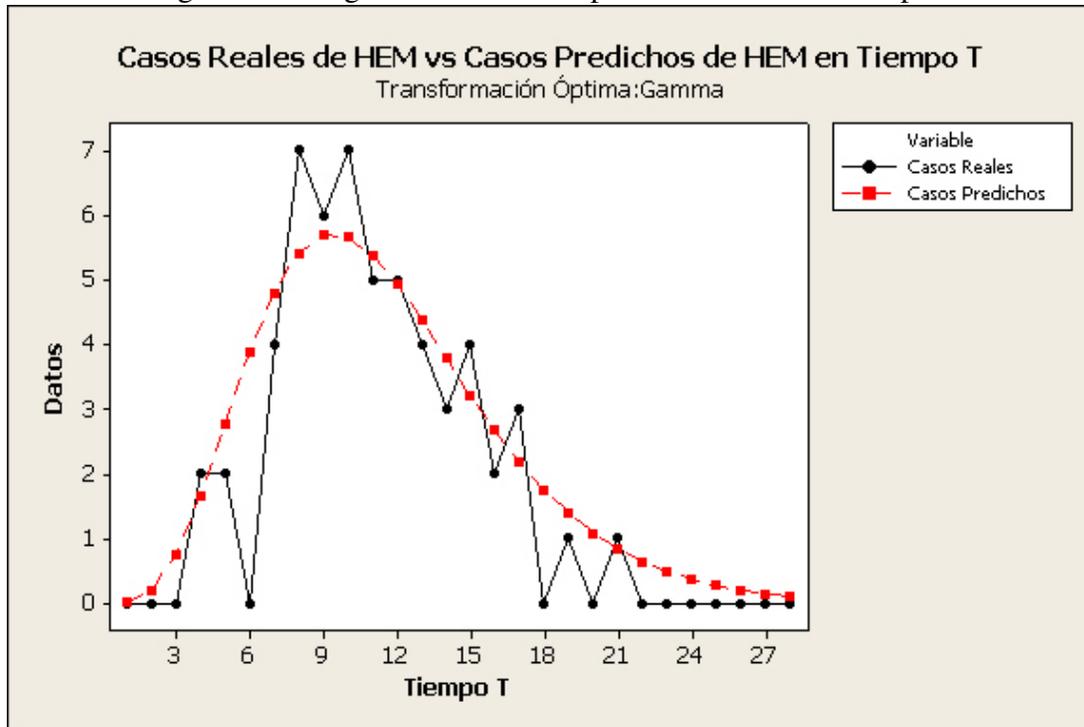
$$\ln(HEM) = -0,521 + 0,404T - 0,0189T^2$$

Predictor	Coef	SE Coef	T	P
constante	-0.5207	0.4500	-1.16	0.270
T	0.40357	0.07884	5.12	0.000
T^2	-0.018880	0.003146	-6.00	0.000

S = 0.298510	R-Sq = 80.7 %	R-Sq(adj) = 77.5 %
PRESS = 2.22873	R-Sq(pred) = 59.80 %	

Fuente	DF	SS	MS	F	P
Regresión	2	4.4751	2.2375	25.11	0.000
Error	12	1.0693	0.0891		
Total	14	5.5444			
T	1	1.2658			
T^2	1	3.2092			

Figura 4.8: Diagnósticos de SIDA por Hemofilia en el tiempo T



Variable de Respuesta: Trans

Predictora: Tiempo T

Modelo Original: Del diagrama de puntos entre Casos Diagnosticados de SIDA de contagiados por Transfusiones de sangre versus Tiempo T, se observa que la gráfica tiene la forma de una distribución de la familia gamma, por lo cual se implantó esa transformación:

El modelo se cita a continuación:

$$Y = \alpha * X^{\beta_1} * e^{\beta_2 X}$$

La transformación a usarse para linealizar la gráfica es el siguiente:

$$Z = \ln(Y)$$

El nuevo modelo, ahora linealizado, es el que se cita a continuación:

$$\ln(Trans) = \ln(\alpha) + \beta_1 * \ln(Year) + \beta_2 * Year$$

Contraste de hipótesis:

$H_o : \beta_i = 0, i = 1, 2$

vs.

$H_1 : \exists \beta_i \neq 0$

La ecuación de regresión obtenida por el programa Minitab, así como el análisis de varianza se presentan a continuación:

$$\ln(Trans) = - 6.29 + 7.01 \ln(T) - 0.705 T$$

Predictor	Coef	SE Coef	T	P
constante	-6.293	1.569	-4.01	0.002
ln(T)	7.015	1.114	6.30	0.000
T	-0.70478	0.09718	-7.25	0.000

S = 0.531288	R-Sq = 83.3 %	R-Sq(adj) = 80.5 %
PRESS = 11.4441	R-Sq(pred) = 43.46 %	

Análisis de Varianza

Fuente	DF	SS	MS	F	P
Regresión	2	16.8527	8.4264	29.85	0.000
Error	12	3.3872	0.2823		
Total	14	20.2399			
$\ln(T)$	1	2.0062			
T	1	14.8465			

La siguiente transformación aplicada a la variable TRANS fue la de la distribución normal y los resultados se muestran a continuación:

$$\ln(Trans) = 0.06 + 0.453 T - 0.0206 T^2$$

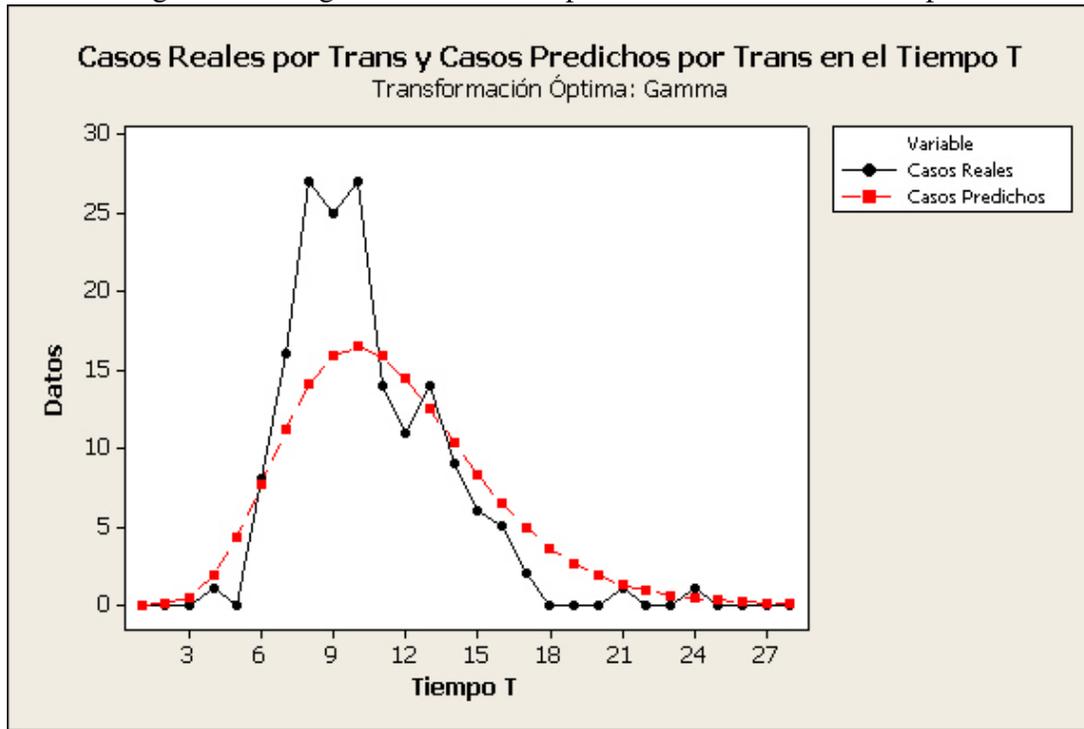
Predictor	Coef	SE Coef	T	P
constante	0.057	1.103	0.05	0.960
T	0.4533	0.1730	2.62	0.022
T^2	-0.020623	0.006128	-3.37	0.006

S = 0.790456	R-Sq = 63.0 %	R-Sq(adj) = 56.8 %
PRESS = 24.8923	R-Sq(pred) = 0.00 %	

Análisis de Varianza

Fuente	DF	SS	MS	F	P
Regresión	2	12.7421	6.3710	10.20	0.003
Error	12	7.4979	0.6248		
Total	14	20.2399			
T	1	5.6659			
T^2	1	7.0761			

Figura 4.9: Diagnósticos de SIDA por Transfusiones en el tiempo T



Variable de Respuesta: TV

Predictora: Tiempo T

Modelo Original:

Al efectuar el diagrama de puntos de T vs TV, observamos una gráfica muy parecida a la curva normal, por lo tanto podemos pensar en el siguiente modelo:

$$Y = \alpha * e^{\beta_1 * X + \beta_2 * X^2}$$

La transformación usada para linealizar la ecuación es el siguiente:

$$Z = \ln(Y)$$

Y el nuevo modelo, ahora linealizado es el que se cita a continuación:

$$\ln(TV) = \ln(\alpha) + \beta_1 T + \beta_2 T^2$$

Contraste de hipótesis:

$$H_0 : \beta_i = 0, i = 1, 2$$

vs.

$$H_1 : \exists \beta_i \neq 0$$

La ecuación de regresión obtenida es la siguiente:

$$\ln(TV) = 0.082 + 0.470 T - 0.0174 T^2$$

Predictor	Coef	SE Coef	T	P
constante	0.0819	0.6747	0.12	0,904
T	0.47003	0.09855	4.77	0.000
T^2	-0.017436	0.003108	-5.61	0.000

S = 0.795432	R-Sq = 65.1 %	R-Sq(adj) = 62.1 %
PRESS = 20.9649	R-Sq(pred) = 49.77 %	

Fuente	DF	SS	MS	F	P
Regresión	2	27.184	13.592	21.48	0.000
Error	23	14.552	0.633		
Total	25	41.736			
T	1	7.265			
T^2	1	19.918			

La siguiente transformación usada fue la de una distribución gamma y los resultados se muestran a continuación:

$$\ln(TV) = - 5.04 - 0.546 T + 6.08 \ln(T)$$

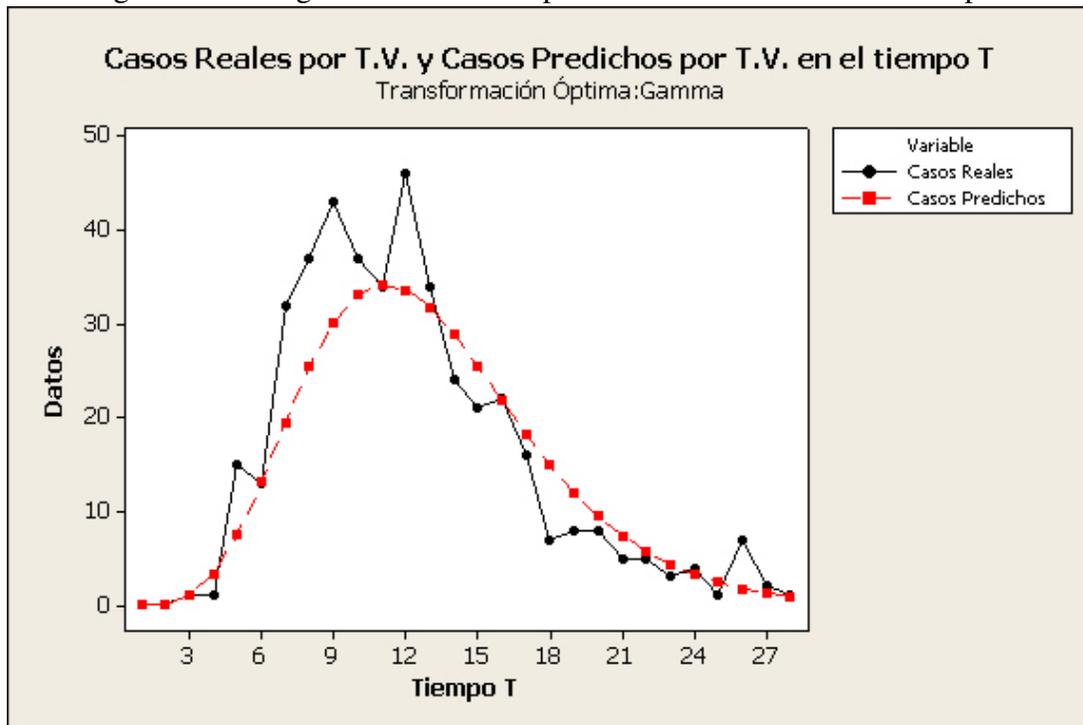
Predictor	Coef	SE Coef	T	P
constante	-5.0437	0.8813	-5.72	0.000
T	-0.54613	0.05037	-10.84	0.000
$\ln(T)$	6.0772	0.6181	9.83	0.000

S = 0.536682	R-Sq = 84.1 %	R-Sq(adj) = 82.7 %
PRESS = 9.20046	R-Sq(pred) = 77.96 %	

Análisis de Varianza

Fuente	DF	SS	MS	F	P
Regresión	2	35.111	17.556	60.95	0.000
Error	23	6.625	0.288		
Total	25	41.736			
T	1	7.265			
$\ln(T)$	1	27.846			

Figura 4.10: Diagnósticos de SIDA por transmisión Vertical en el tiempo T



OTRAS VARIABLES

Para el caso pediátrico se observó que el diagrama de puntos de la variable tiempo T versus las variables predictoras RBI, TPED, HPED y RBIPED no tenia características especiales. Por tal motivo esas variables no fueron analizadas individualmente, y solo fueron consideradas en el análisis de regresión. A continuación se muestran los diagramas de puntos de los casos mencionados:

Figura 4.11: Diagnósticos de SIDA por causas bajo investigación en el tiempo T

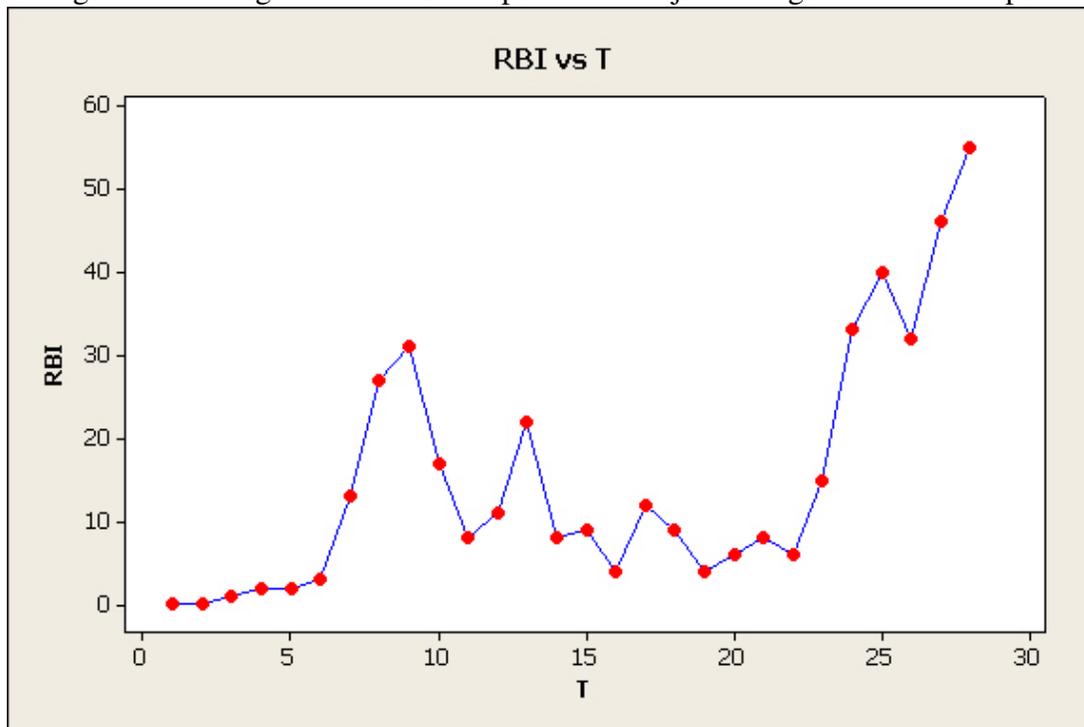


Figura 4.12: Diagnósticos de SIDA por transfusiones casos pediátricos en el tiempo T

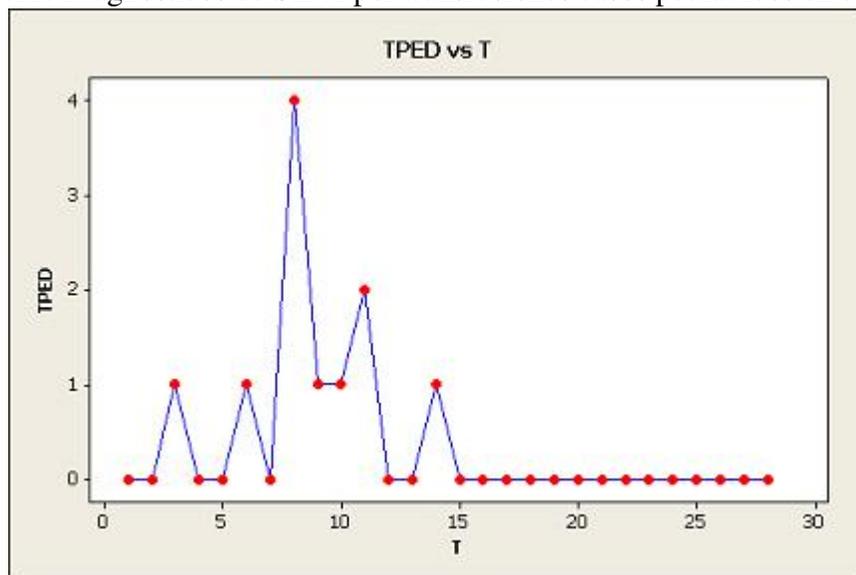
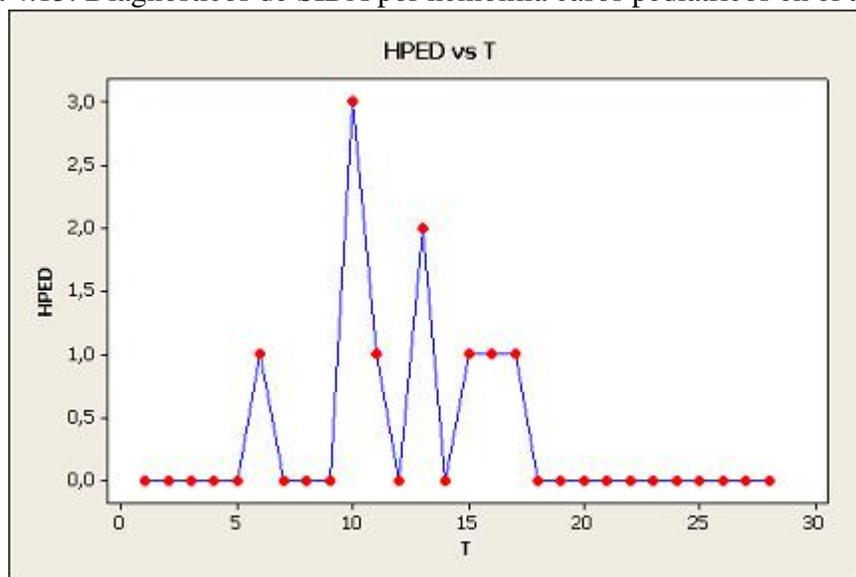


Figura 4.13: Diagnósticos de SIDA por hemofilia casos pediátricos en el tiempo T



4.7.2. Género masculino

Dado el comportamiento de las predictoras para los casos de muertes de hombres por SIDA, se desarrollaron múltiples transformaciones sobre las mismas de tal forma que se pueda determinar

el tipo de distribución y el peso que tiene cada una sobre la variable de respuesta.

Variable de Respuesta: Muertes

Predictora: Tiempo T

La primera transformación aplicada a la variable fue la de la distribución normal, y el modelo es el siguiente:

$$\ln(\text{Muertes}) = \ln(\alpha) + \beta_1 T + \beta_2 T^2$$

Contraste de hipótesis:

$$H_0 : \beta_i = 0$$

vs.

$$H_1 : \exists \beta_i \neq 0$$

La ecuación de regresión obtenida es la siguiente:

$$\ln(\text{Muertes}) = 0.879 + 0.779 T - 0.0235 T^2$$

Predictor	Coef	SE Coef	T	P
Predictor	Coef	SE Coef	T	P
constante	0.8791	0.4405	2.00	0.057
T	0.77883	0.07001	11.12	0.000
T2	-0.023508	0.002343	-10.03	0.000

S = 0.722147	R-Sq = 84.2 %	R-Sq(adj) = 83.0 %
PRESS = 18.4256	R-Sq(pred) = 77.74 %	

Fuente	DF	SS	MS	F	P
Regresión	2	69.724	34.862	66.85	0.000
Error	25	13.037	0.521		
Total	27	82.761			
T	1	17.221			
T ²	1	52.503			

Sin embargo, también se puede hacer una transformación a una distribución gamma, siendo los resultados obtenidos los que se citan a continuación:

$$\ln(\text{Muertes}) = - 0.609 - 0.344 T + 4.66 \ln(T)$$

Predictor	Coef	SE Coef	T	P
constante	-0.6093	0.2635	-2.31	0.029
T	-0.34386	0.02181	-15.77	0.000
ln(T)	4.6567	0.2124	21.92	0.000

S = 0.360047	R-Sq = 96.1 %	R-Sq(adj) = 95.8 %
PRESS = 8.65485	R-Sq(pred) = 89.54 %	

Fuente	DF	SS	MS	F	P
Regresión	2	79.520	39.760	306.71	0.000
Error	25	3.241	0.130		
Total	27	82.761			
T	1	17.221			
ln(T)	1	62.300			

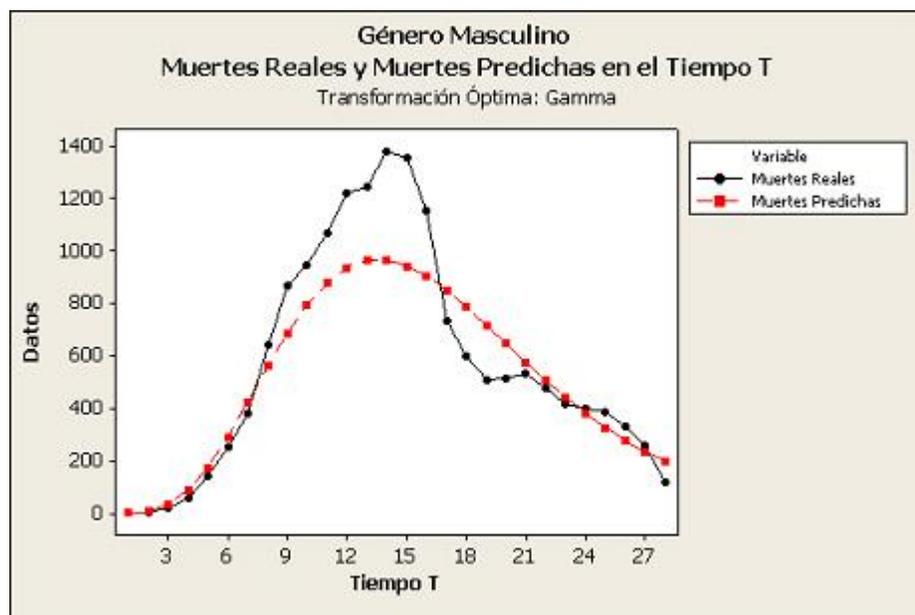


Figura 4.14: Muertes por SIDA en el tiempo T. Género Masculino

Variable de Respuesta: Casos

Predictora: Tiempo T

La primera transformación usada para la variable predictora CASOS fue la distribución normal.

$$\begin{aligned} \ln(Y) &= \ln(\alpha) + \beta_1 X + \beta_2 X^2 \\ Z &= \ln(\alpha) + \beta_1 X + \beta_2 X^2 \\ \ln(Casos) &= \ln(\alpha) + \beta_1 T + \beta_2 T^2 \end{aligned}$$

Contraste de hipótesis:

$$H_o : \beta_i = 0$$

vs.

$$H_1 : \exists \beta_i \neq 0$$

El modelo de regresión obtenido fue el siguiente:

$$\ln(Casos) = 1.58 + 0.738 T - 0.0221 T^2$$

Predictor	Coef	SE Coef	T	P
constante	1.5822	0.4520	3.50	0.002
T	0.73753	0.07185	10.27	0.000
T^2	-0.022064	0.002404	-9.18	0.000

S = 0.741051	R-Sq = 82.3 %	R-Sq(adj) = 80.8 %
PRESS = 20.7574	R-Sq(pred) = 73.19 %	

Fuente	DF	SS	MS	F	P
Regresión	2	63.681	31.840	57.98	0,000
Error	25	13.729	0.549		
Total	27	77.410			
T	1	17.432			
T^2	1	46.249			

La transformación a una gamma para esta variable obtuvo los siguientes resultados:

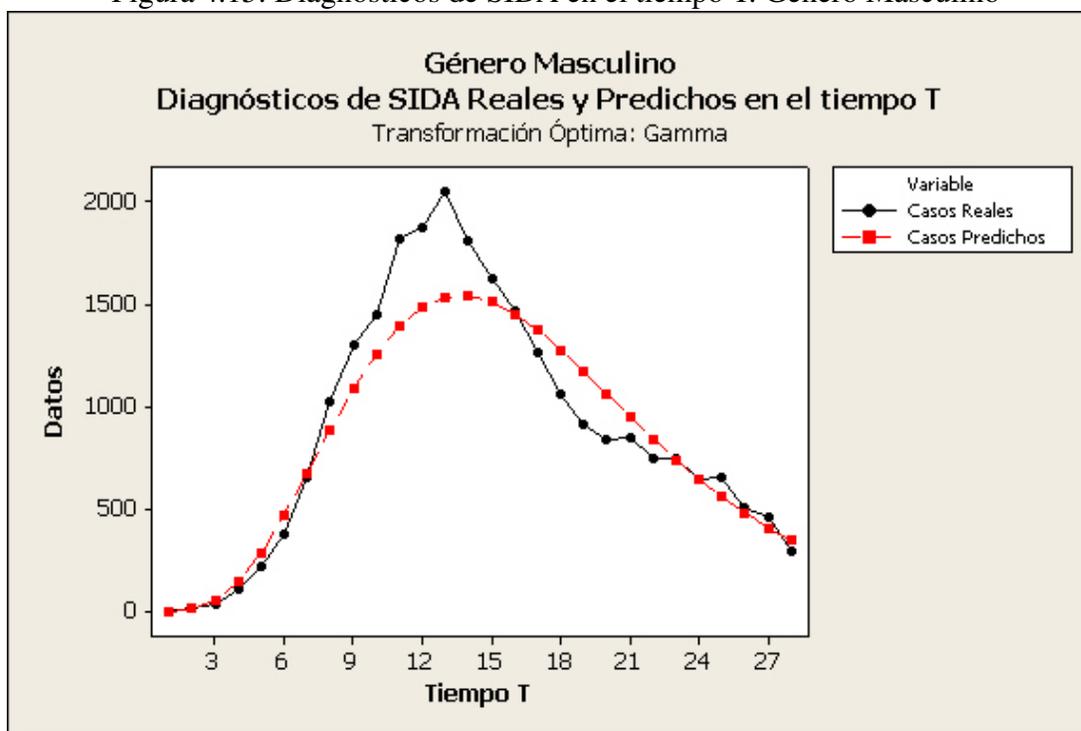
$$\ln Casos = 0.019 - 0.331T + 4.53 \ln(T)$$

Predictor	Coef	SE Coef	T	P
constante	0.0193	0.1509	0.13	0.009
T	-0.33112	0.01249	-26.51	0.000
ln(T)	4.5284	0.1217	37.22	0.000

S = 0.206209	R-Sq = 98.6 %	R-Sq(adj) = 98.5 %
PRESS = 1.75835	R-Sq(pred) = 97.73 %	

Fuente	DF	SS	MS	F	P
Regresión	2	76.347	38.173	897.73	0.000
Error	25	1.063	0.043		
Total	27	77.410			
T	1	17.432			
ln(T)	1	58.915			

Figura 4.15: Diagnósticos de SIDA en el tiempo T. Género Masculino



Variable de Respuesta: Contagio por Contacto Homosexual y Bisexual
Predictora: Tiempo T

$$\begin{aligned} \ln(Y) &= \ln(\alpha) + \beta_1 X + \beta_2 X^2 \\ Z &= \ln(\alpha) + \beta_1 X + \beta_2 X^2 \\ n(HSH) &= \ln(\alpha) + \beta_1 T + \beta_2 T^2 \end{aligned}$$

Contraste de hipótesis:

$$H_o : \beta_i = 0$$

vs.

$$H_1 : \exists \beta_i \neq 0$$

La ecuación de regresión es la siguiente:

$$\ln(HSH) = 1.04 + 0.601 T - 0.0180 T^2$$

Predictor	Coef	SE Coef	T	P
constante	1.0433	0.3432	3.04	0.005
T	0.60062	0.05454	11.01	0.000
T^2	-0.017980	0.001825	-9.85	0.000

S = 0.562549	R-Sq = 84.2 %	R-Sq(adj) = 82.9 %
PRESS = 11.8311	R-Sq(pred) = 76.38 %	

Fuente	DF	SS	MS	F	P
Regresión	2	42.175	21.087	66.64	0.000
Error	25	7.912	0.316		
Total	27	50.087			
T	1	11.463			
T^2	1	30.712			

Luego aplicamos la transformación a la distribución gamma a la variable predictora, siendo los resultados los que se presentan a continuación:

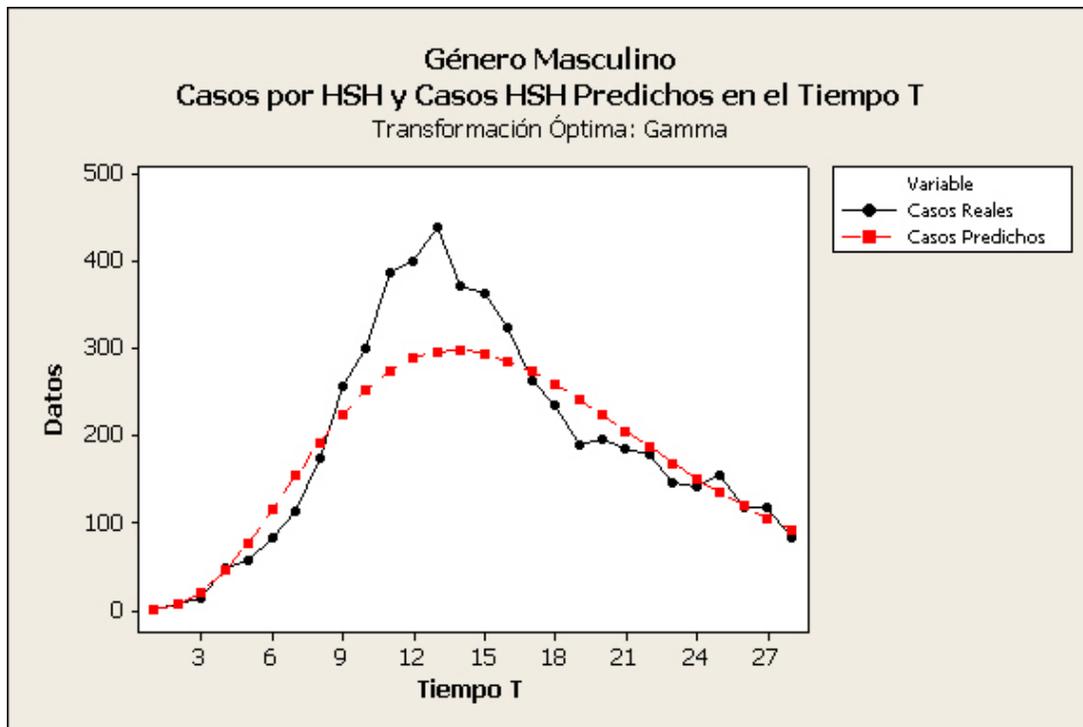
$$\ln(HSH) = -0.135 - 0.262 T + 3.60 \ln(T)$$

Predictor	Coef	SE Coef	T	P
constante	-0.1350	0.1730	-0.78	0.042
T	-0.26164	0.01432	-18.27	0,000
$\ln(T)$	3.5996	0.1395	25.81	0,000

S = 0.236420	R-Sq = 97.2 %	R-Sq(adj) = 97.0 %
PRESS = 2.43548	R-Sq(pred) = 95.14 %	

Fuente	DF	SS	MS	F	P
Regresión	2	48.689	24.345	435.55	0.000
Error	25	1.397	0.056		
Total	27	50.087			
T	1	11.463			
ln(T)	1	37.226			

Figura 4.16: Diagnósticos de SIDA por Contacto Homosexual en el tiempo T. Género Masculino



Variable de Respuesta: Contagio por Uso de Drogas Intravenosas

Predictora: Tiempo T

La siguiente variable a la cual se le aplicaron transformaciones fue la UDI en género masculino. La primera transformación a usarse fue la de la distribución normal, la cual se explica a continuación:

$$\begin{aligned} \ln(Y) &= \ln(\alpha) + \beta_1 X + \beta_2 X^2 \\ Z &= \ln(\alpha) + \beta_1 X + \beta_2 X^2 \\ n(UDI) &= \ln(\alpha) + \beta_1 T + \beta_2 T^2 \end{aligned}$$

Contraste de hipótesis:

$$H_0 : \beta_i = 0$$

vs.

$$H_1 : \exists \beta_i \neq 0$$

La ecuación de regresión es:

$$\ln(UDI) = 0.794 + 0.572 T - 0.0187 T^2$$

Predictor	Coef	SE Coef	T	P
constante	0.7944	0.4690	1.69	0.103
T	0.57204	0.07150	8.00	0.000
T^2	-0.018680	0.002322	-8.05	0.000

S = 0.653247	R-Sq = 73.1 %	R-Sq(adj) = 70.9 %
PRESS = 15.6849	R-Sq(pred) = 58.82 %	

Fuente	DF	SS	MS	F	P
Regresión	2	27.847	13.924	32.63	0.000
Error	24	10.242	0.427		
Total	26	38.089			
T	1	0.223			
T^2	1	27.624			

Los resultados para una transformación a una distribución gamma es la que se da a continuación:

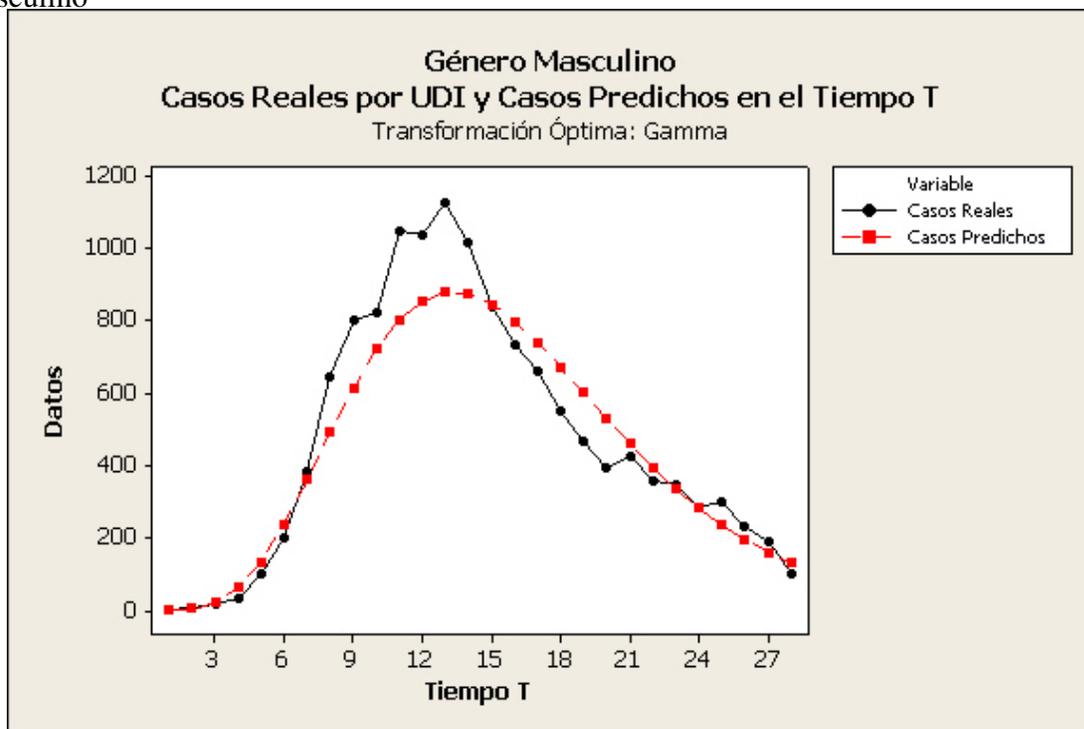
$$\ln(UDI) = -1,67 - 0,400T + 5,32\ln(T)$$

Predictor	Coef	SE Coef	T	P
constante	-1.6652	0.3034	-5.49	0.000
T	-0.40046	0.02041	-19.63	0.000
$\ln(T)$	5.3220	0.2277	23.38	0.000

S = 0.264826	R-Sq = 96.2 %	R-Sq(adj) = 95.9 %
PRESS = 3.15908	R-Sq(pred) = 92.87 %	

Fuente	DF	SS	MS	F	P
Regresión	2	42.640	21.320	303.99	0.000
Error	24	1.683	0.070		
Total	26	44.323			
T	1	4.317			
$\ln(T)$	1	38.322			

Figura 4.17: Diagnósticos de SIDA por Uso de Drogas Intravenosas en el tiempo T. Género Masculino



Variable de Respuesta: Contagio por contacto homosexual y bisexual y uso de drogas intravenosas

Predictora: Tiempo T

La primera transformación aplicada a esta variable fue la de la distribución normal. A continuación se muestra el modelo transformado:

$$\ln(HSHUDI) = \ln(\alpha) + \beta_1 T + \beta_2 T^2$$

Contraste de hipótesis:

$$H_o : \beta_i = 0$$

vs.

$$H_1 : \exists \beta_i \neq 0$$

La ecuación de regresión obtenida fue la siguiente :

$$\ln(HSUUDI) = 0.794 + 0.572 T - 0.0187 T^2$$

Predictor	Coef	SE Coef	T	P
constante	0.7944	0.4690	1.69	0.103
T	0.57204	0.07150	8.00	0.000
T ²	-0.018680	0.002322	-8.05	0.000

S = 0.653247	R-Sq = 73.1 %	R-Sq(adj) = 70.9 %
PRESS = 15.6849	R-Sq(pred) = 58.82 %	

Fuente	DF	SS	MS	F	P
Regresión	2	27.847	13.924	32.63	0.000
Error	24	10.242	0.427		
Total	26	38.089			
T	1	0.223			
T ²	1	27.624			

Los resultados de la transformación gamma para la variable HSHUDI es la que se detalla a continuación:

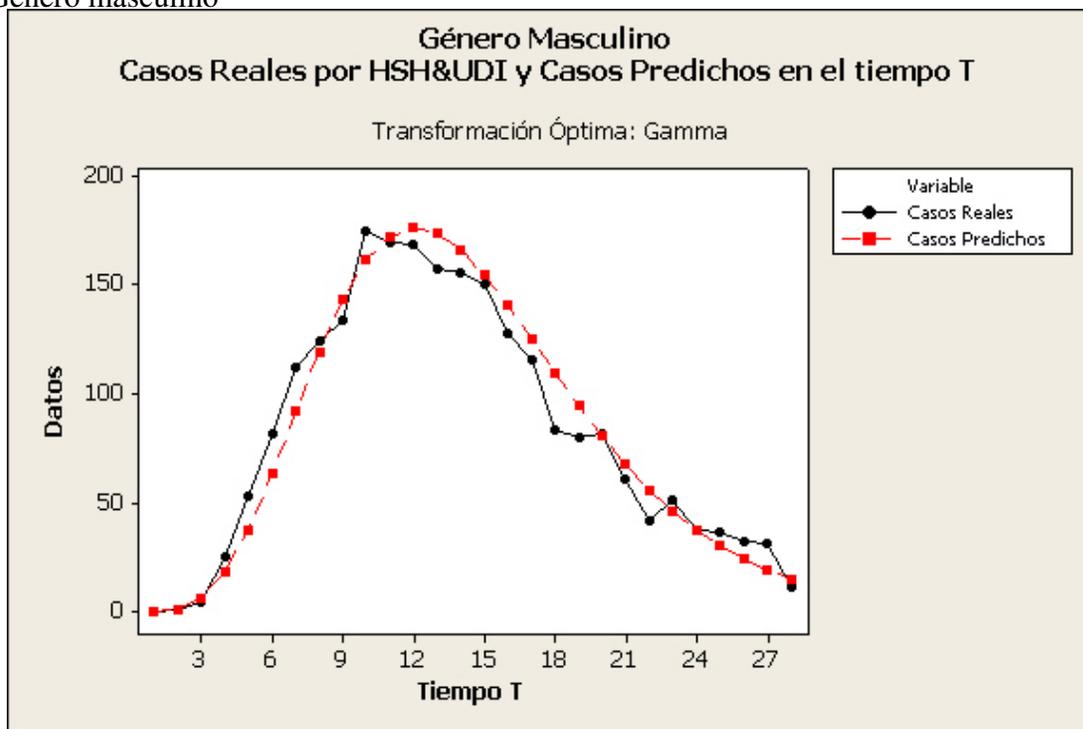
$$\ln(HSUUDI) = -2.59 - 0.430 T + 5.20 \ln(T)$$

Predictor	Coef	SE Coef	T	P
constante	-2.5909	0.2627	-9.86	0.000
T	-0.42990	0.01766	-24.34	0.000
ln(T)	5.2013	0.1971	26.39	0.000

S = 0.229245	R-Sq = 96.7 %	R-Sq(adj) = 96.4 %
PRESS = 1.80488	R-Sq(pred) = 95.26 %	

Fuente	DF	SS	MS	F	P
Regresión	2	36.827	18.414	350.38	0.000
Error	24	1.261	0.053		
Total	26	38.089			
T	1	0.223			
$\ln(T)$	1	36.605			

Figura 4.18: Diagnósticos por contacto homosexual y uso de drogas intravenosas en el tiempo T. Género masculino



Variable de Respuesta: Contagio por contacto homosexual y bisexual

Predictora: Tiempo T

La primera transformación a usarse para esta variable fue la de la distribución normal, siendo los resultados los que se muestran a continuación:

$$\ln(CH) = \ln(\alpha) + \beta_1 T + \beta_2 T^2$$

Contraste de hipótesis:

$$H_o : \beta_i = 0$$

vs.

$$H_1 : \exists \beta_i \neq 0$$

La ecuación de regresión obtenida en minitab fue la siguiente:

$$\ln(CH) = - 1.92 + 0.822 T - 0.0221 T^2$$

Predictor	Coef	SE Coef	T	P
constante	-1.9201	0.7157	-2.68	0.014
T	0.82236	0.09225	8.91	0.000
T^2	-0.022149	0.002673	-8.29	0.000

S = 0.503100	R-Sq = 81.5 %	R-Sq(adj) = 79.6 %
PRESS = 7.64154	R-Sq(pred) = 72.01 %	

Fuente	DF	SS	MS	F	P
Regresión	2	22.237	11.119	43.93	0.000
Error	20	5.062	0.253		
Total	22	27.299			
T	1	4.862			
T^2	1	17.376			

La segunda transformación presentó los siguientes resultados:

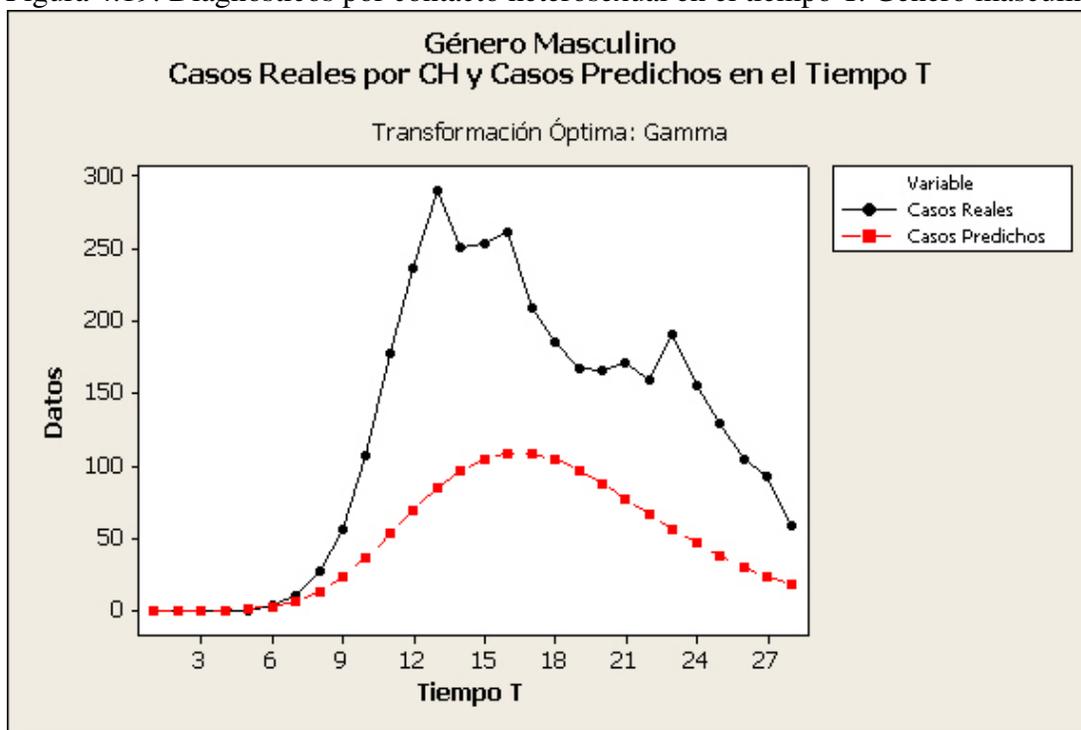
$$\ln(CH) = - 14.0 - 0.634 T + 10.7 \ln(T)$$

Predictor	Coef	SE Coef	T	P
constante	-13.955	1.073	-13.01	0.000
T	-0.63414	0.04363	-14.54	0.000
$\ln(T)$	10.7231	0.6516	16.46	0.000

S = 0.277758	R-Sq = 94.3 %	R-Sq(adj) = 93.8 %
PRESS = 2.16469	R-Sq(pred) = 92.07 %	

Fuente	DF	SS	MS	F	P
Regresión	2	25.756	12.878	166.93	0.000
Error	20	1.543	0.077		
Total	22	27.299			
T	1	4.862			
$\ln(T)$	1	20.895			

Figura 4.19: Diagnósticos por contacto heterosexual en el tiempo T. Género masculino



Variable de Respuesta: Contagio por hemofilia

Predictora: Tiempo T

Para la variable predictora Hemofilia la transformación a una distribución gamma generó los siguientes resultados:

$$\ln(HEM) = - 3.40 - 0.435T + 4.06 \ln(T)$$

Predictor	Coef	SE Coef	T	P
constante	-3.3962	0.7257	-4.68	0.001
T	-0.43547	0.05326	-8.18	0.000
$\ln(T)$	4.0560	0.5562	7.29	0.000

S = 0.236072	R-Sq = 86.1 %	R-Sq(adj) = 83.8 %
PRESS = 1.28636	R-Sq(pred) = 73.32 %	

Fuente	DF	SS	MS	F	P
Regresión	2	4.1520	2.0760	37.25	0.000
Error	12	0.6688	0.0557		
Total	14	4.8208			
T	1	1.1885			
$\ln(T)$	1	2.9635			

La transformación a distribución normal para HEM obtuvo los siguientes resultados:

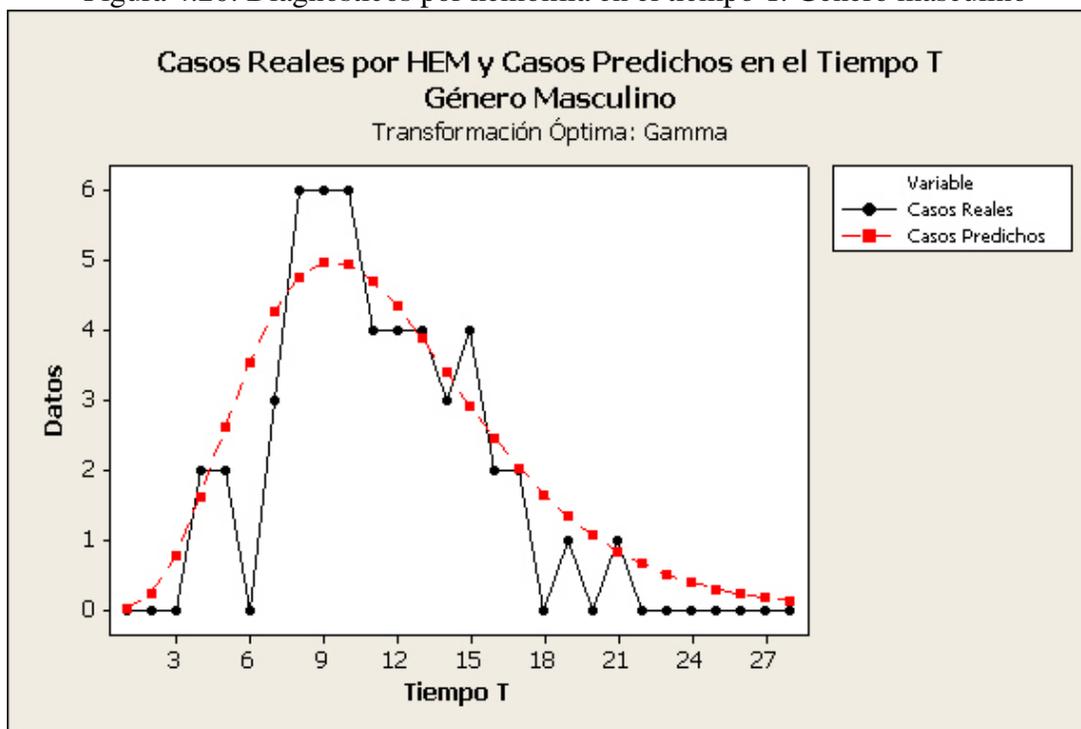
$$\ln HEM = - 0.455 + 0.370 T - 0.0174 T^2$$

Predictor	Coef	SE Coef	T	P
constante	-0.4551	0.4124	-1.10	0.291
T	0.36978	0.07224	5.12	0.000
T2	-0.017427	0.002883	-6.05	0.000

S = 0.273520	R-Sq = 81.4 %	R-Sq(adj) = 78.32 %
PRESS = 1.95512	R-Sq(pred) = 59.44 %	

Fuente	DF	SS	MS	F	P
Regresión	2	3.9230	1.9615	26.22	0.000
Error	12	0.8978	0.0748		
Total	14	4.8208			
T	1	1.1885			
T^2	1	2.7345			

Figura 4.20: Diagnósticos por hemofilia en el tiempo T. Género masculino



Variable de Respuesta: Contagio por transfusiones

Predictora: Tiempo T

La primera transformación aplicada a la variable TRANS fue la distribución gamma, en la cual se obtuvieron los siguientes resultados:

$$\ln(TRANS) = -3.80 - 0.457 T + 4.47 \ln(T)$$

Predictor	Coef	SE Coef	T	P
constante	-3.799	2.085	-1.82	0.098
T	-0.4569	0.1353	-3.38	0.007
$\ln(T)$	4.467	1.500	2.98	0.014

S = 0.696271	R-Sq = 55.0 %	R-Sq(adj) = 46.0 %
PRESS = 22.4752	R-Sq(pred) = 0.00 %	

Fuente	DF	SS	MS	F	P
Regresión	2	5.9296	2.9648	6.12	0.018
Error	10	4.8479	0.4848		
Total	12	10.7775			
T	1	1.6322			
$\ln(T)$ =	1	4.2974			

La transformación a normal para la variable es la que sigue a continuación:

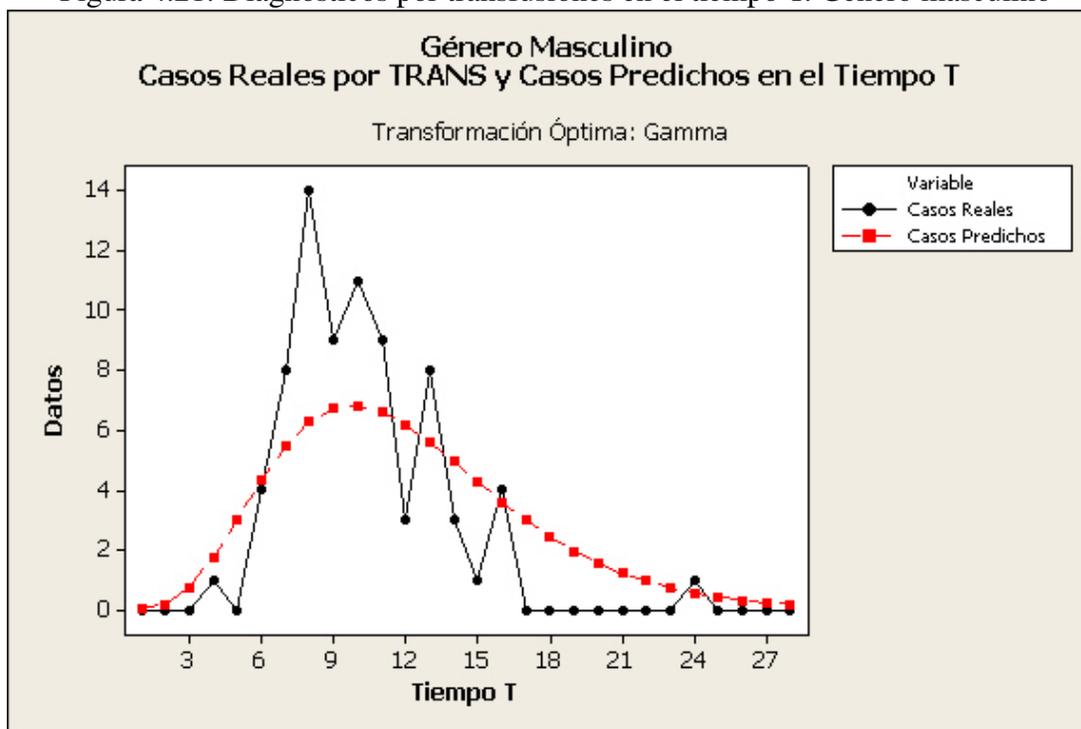
$$\ln(TRANS) = 0.35 + 0.260 T - 0.0121 T^2$$

Predictor	Coef	SE Coef	T	P
constante	0.350	1.176	0.30	0.772
T	0.2596	0.1852	1.40	0.191
T^2	-0.012132	0.006589	-1.84	0.095

S = 0.826447	R-Sq = 36.6 %	R-Sq(adj) = 24.0 %
PRESS = 75.2837	R-Sq(pred) = 0.00 %	

Fuente	DF	SS	MS	F	P
Regresión	2	3.9474	1.9737	2.89	0.102
Error	10	6.8301	0.6830		
Total	12	10.7775			
T	1	1.6322			
T^2	1	2.3152			

Figura 4.21: Diagnósticos por transfusiones en el tiempo T. Género masculino



Se observa que el coeficiente de determinación es muy bajo por lo cual se concluye que no se determinó un modelo de buen ajuste para esta variable.

Variable de Respuesta: transmisión Vertical

Predictora: Tiempo T

Para la variable transmisión vertical se usaron ambas transformaciones, y la primera transformación aplicada fue la normal.

$$\ln(TV) = \ln(\alpha) + \beta_1 T + \beta_2 T^2$$

Contraste de hipótesis:

$$H_o : \beta_i = 0$$

vs.

$$H_1 : \exists \beta_i \neq 0$$

La ecuación de regresión obtenida es la siguiente:

$$\ln(TV) = - 0.006 + 0.377 T - 0.0146 T^2$$

Predictor	Coef	SE Coef	T	P
constante	-0.0057	0.6825	-0.01	0.993
T	0.3774	0.1014	3.72	0.001
T^2	-0.014616	0.003267	-4.47	0.000

S = 0.792832	R-Sq = 57.0 %	R-Sq(adj) = 53.0 %
PRESS = 19.7548	R-Sq(pred) = 38.52 %	

Fuente	DF	SS	MS	F	P
Regresión	2	18.3021	9.1511	14.56	0.000
Error	22	13.8288	0.6286		
Total	24	32.1309			
T	1	5.7209			
T^2	1	12.5812			

La transformación a una gamma para TV es la que sigue a continuación:

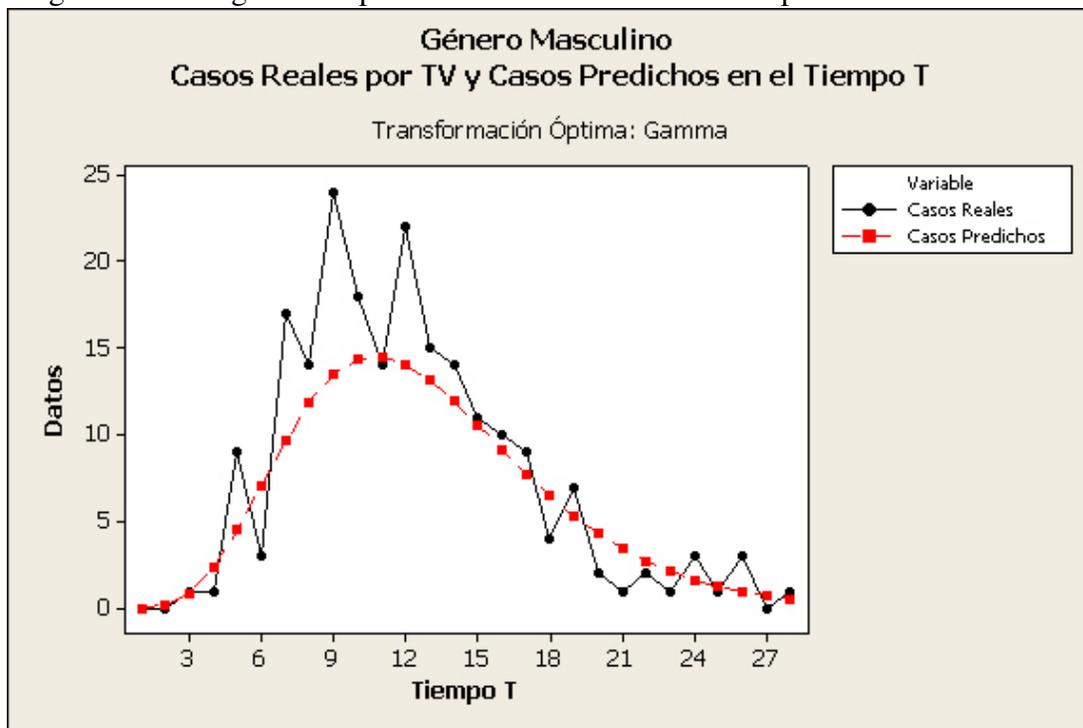
$$\ln(T)V = - 4.28 + 5.06 \ln(T) - 0.471 T$$

Predictor	Coef	SE Coef	T	P
constante	-4.277	1.013	-4,22	0.000
$\ln(T)$	5.0650	0.7193	7,04	0.000
T	-0.47077	0,05989	-7.86	0.000

S = 0.60739032	R-Sq = 74.7 %	R-Sq(adj) = 72.4 %
PRESS = 10.7531	R-Sq(pred) = 66.53 %	

Fuente	DF	SS	MS	F	P
Regresión	2	24.015	12.007	32.55	0.000
Error	22	8.116	0.369		
Total	24	32.131			
ln(T)	1	1.222			
T	1	22.792			

Figura 4.22: Diagnósticos por trasmisión Vertical en el tiempo T. Género Masculino



OTRAS VARIABLES

Al igual que en el caso total, para el género masculino hay tipos de contagio de SIDA que han sido muy raros, o que no han sido determinados. Estas variables se llaman en nuestro estudio RBI (riesgo bajo investigación), TPED (transfusiones en casos pediátricos), HPED (hemofilia en casos pediátricos) y RBIPED (riesgo bajo investigación en casos pediátricos).

Como estas variables no cuentan con mucha información, para las mismas no se efectuó transformaciones. Sin embargo si fue de interés observar su comportamiento en el tiempo, por lo cual se graficaron diagramas de puntos. A continuación los diagramas de puntos de estas variables.

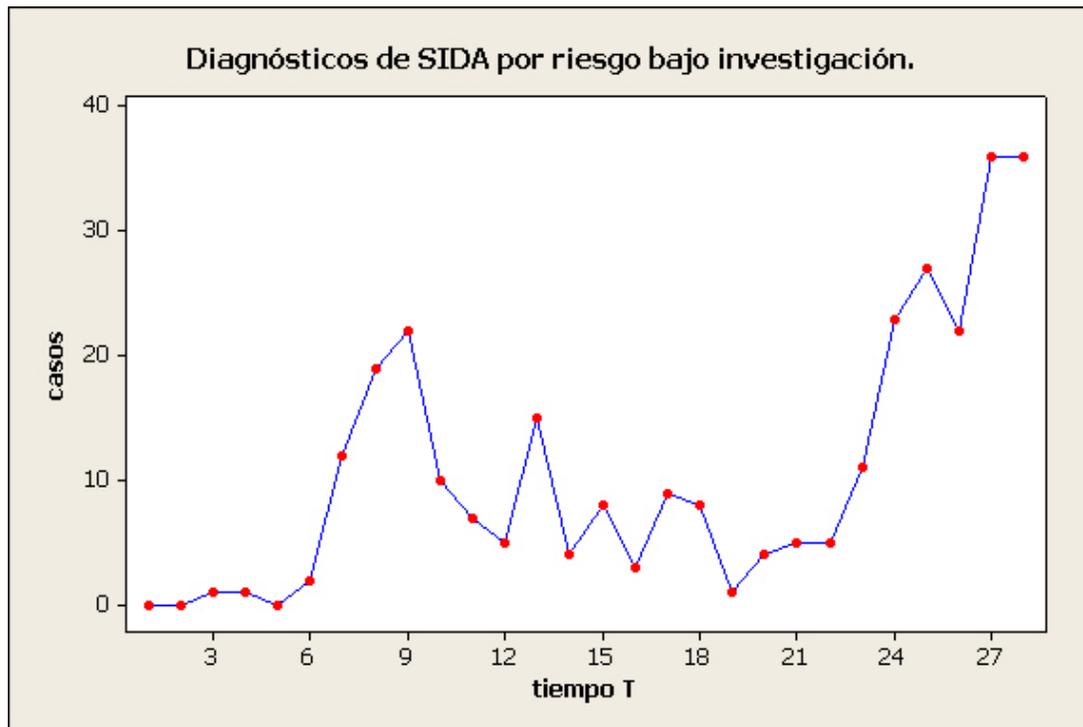


Figura 4.23: Diagnósticos de SIDA por riesgo bajo investigación en el tiempo T. Género Masculino

Figura 4.24: Diagnósticos por transfusiones pediátricas en el tiempo T. Género Masculino

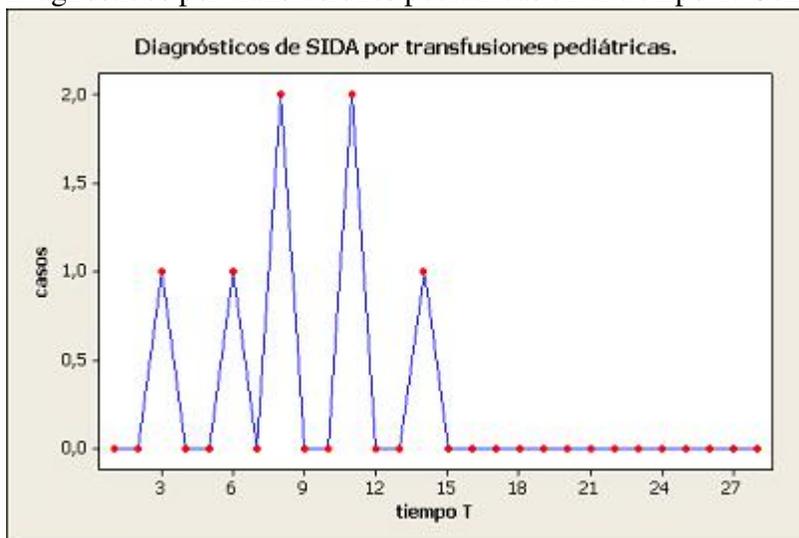


Figura 4.25: Diagnósticos por hemofilia en casos pediátricos en el tiempo T. Género Masculino

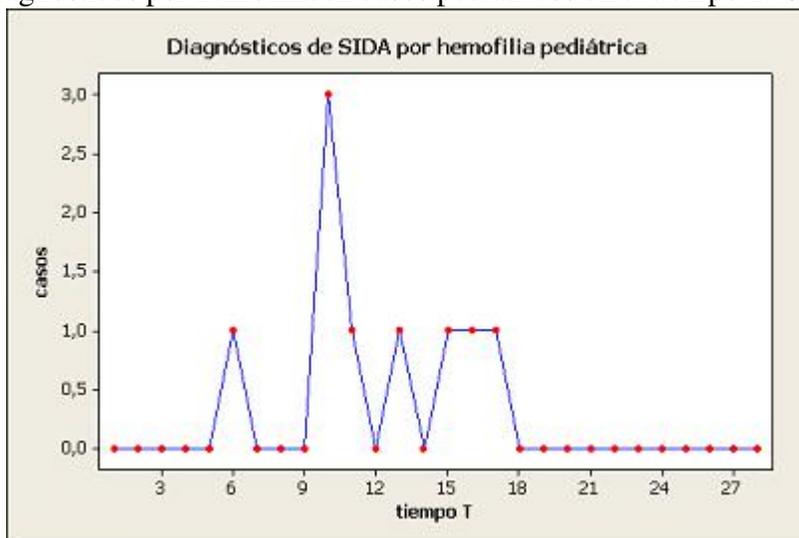
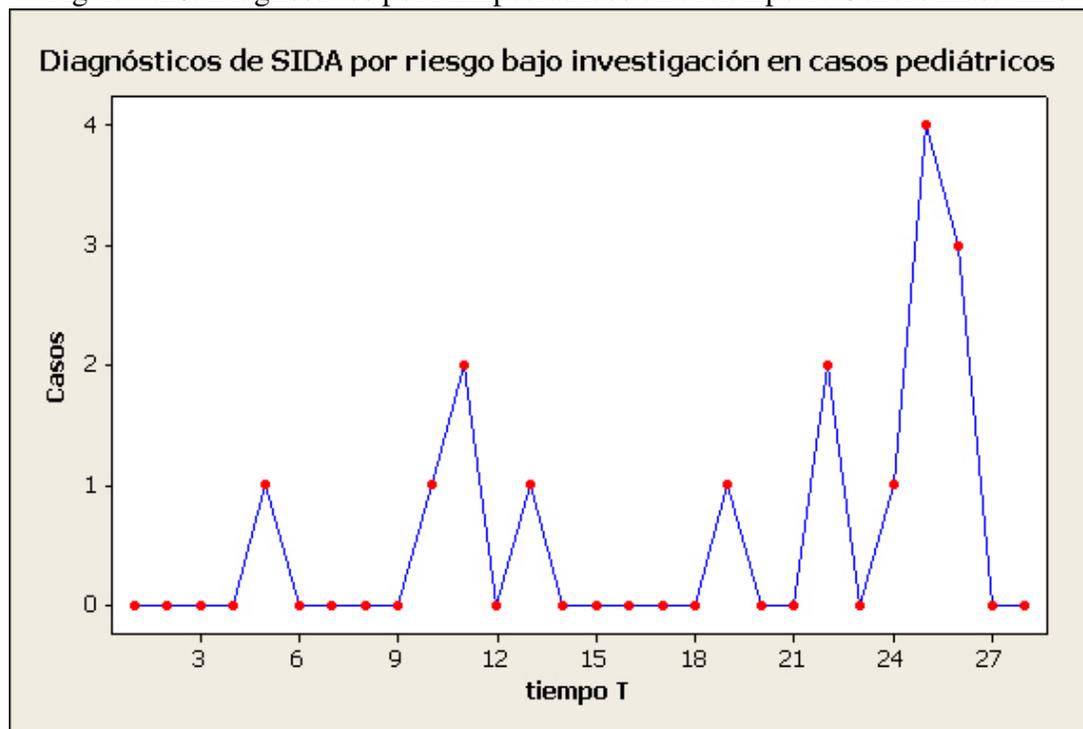


Figura 4.26: Diagnósticos por RBI pediátricos en el tiempo T. Género masculino



4.7.3. Género femenino

Al igual que en el casos anteriores, fue de interés investigar el tipo de comportamiento de cada variable en función del tiempo. Para esto usamos las transformaciones que se detallaron en las secciones anteriores, tanto para el Caso total como para el caso particular del género masculino.

Variable de Respuesta: Muertes

Predictora: Tiempo

Dada la tendencia de esta variable se van a usar las dos transformaciones que se han trabajado anteriormente, la normal y la gamma. La primera transformación que se realizó fue la de la distribución normal, y los resultados se observan a continuación:

$$\ln(\text{Muertes}) = -0.157 + 0.711T - 0.0207 T^2$$

Predictor	Coef	SE Coef	T	P ϖ
constante	-0.1568	0.3697	-0.42	0.675
T	0.71100	0.05636	12.62	0.000
T ²	-0.020744	0.001830	-11.33	0.000

S = 0.514947	R-Sq = 88.1 %	R-Sq(adj) = 87.1 %
PRESS = 8.50105	R-Sq(pred) = 84.05 %	

Fuente	DF	SS	MS	F	P
Regresión	2	46.949	23.475	88.53	0.000
Error	24	6.364	0.265		
Total	26	53.314			
T	1	12.882			
T^2	1	34.067			

Luego, se realizó la transformación del modelo lineal a una distribución gamma, siendo los resultados los siguientes:

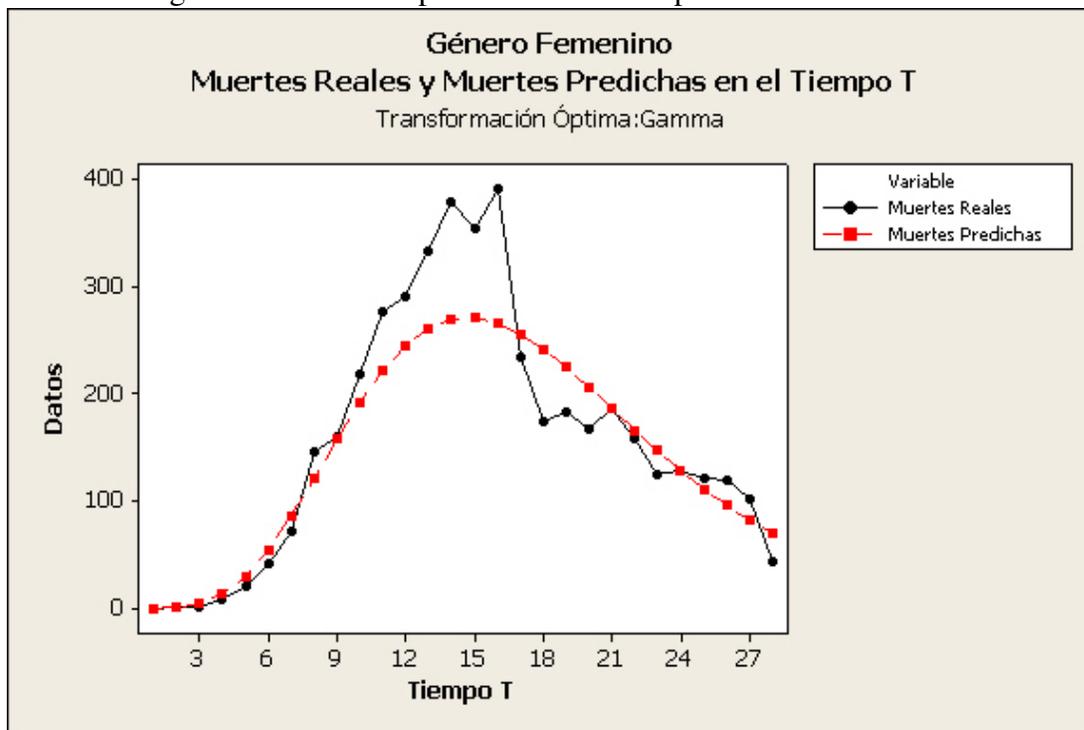
$$\ln(\text{Muertes}) = - 3.26 - 0.357 T + 5.25 \ln(T)$$

Predictor	Coef	SE Coef	T	P
constante	-3.2578	0.4184	-7.79	0.000
T	-0.35663	0.02814	-12.67	0.000
$\ln(T)$	5.2456	0.3139	16.71	0.000

S = 0.365186	R-Sq = 94 %	R-Sq(adj) = 93.5 %
PRESS = 7.57886	R-Sq(pred) = 85.78 %	

Fuente	DF	SS	MS	F	P
Regresión	2	50.113	25.056	187.88	0.000
Error	24	3.201	0.133		
Total	26	53.314			
T	1	12.882			
$\ln(T)$	1	37.230			

Figura 4.27: Muertes por SIDA en el tiempo T. Género femenino



Variable de Respuesta: Casos

Predictora: Tiempo

La primera transformación que se realizó fue la de la distribución normal, y los resultados se observan a continuación:

Transformación normal

La ecuación de regresión obtenida fue:

$$\ln(Casos) = 0.072 + 0.748 T - 0.0214 T^2$$

Predictor	Coef	SE Coef	T	P
constante	0.0722	0.3467	0.21	0.837
T	0.74782	0.05510	13.57	0.000
T^2	-0.021400	0.001844	-11.61	0.000

S = 0.568284	R-Sq = 90.1 %	R-Sq(adj) = 89.3 %
PRESS = 10.7037	R-Sq(pred) = 86.81 %	

Fuente	DF	SS	MS	F	P
Regresión	2	73.078	36.539	113.14	0.000
Error	25	8.074	0.323		
Total	27	81.152			
T	1	29.570			
T^2	1	43.509			

Transformación gamma

La siguiente transformación a usarse fue a la distribución gamma, y los resultados se muestran a continuación:

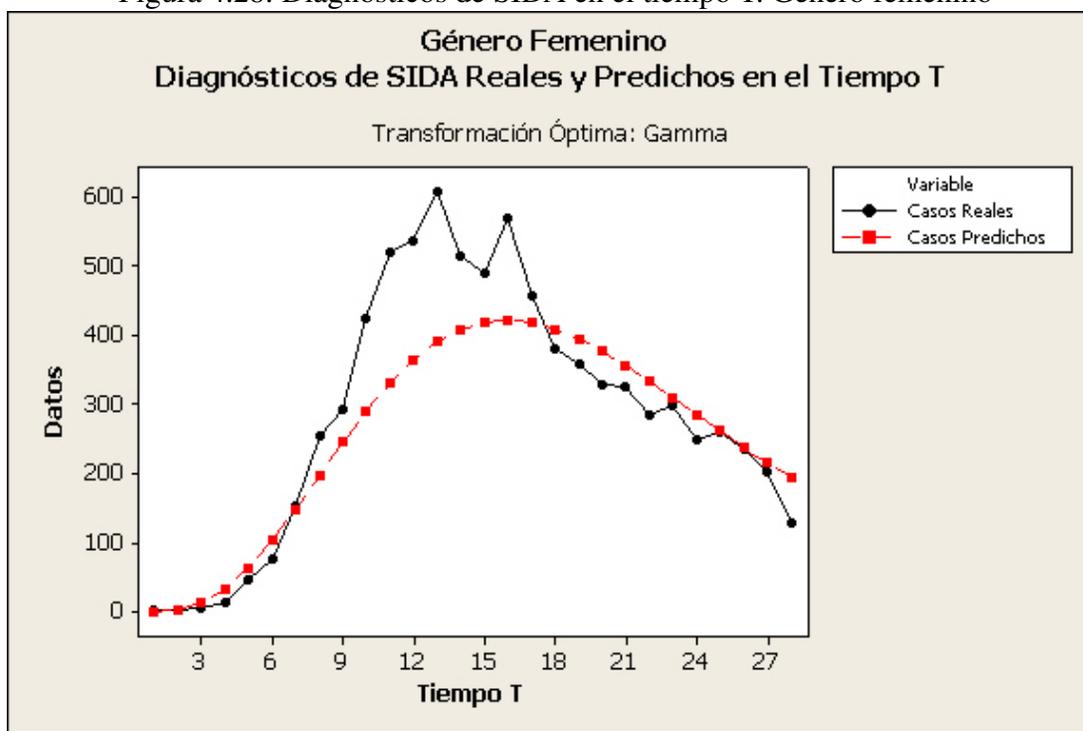
$$\ln(Casos) = -1.04 - 0.252 T + 4.01 \ln(T)$$

Predictor	Coef	SE Coef	T	P
constante	-1.0385	0.3420	-3.04	0.006
T	-0.25219	0.02830	-8.91	0.000
$\ln(T)$	4.0069	0.2756	14.54	0.000

S = 0.467198	R-Sq = 93.3 %	R-Sq(adj) = 92.7 %
PRESS = 14.8734	R-Sq(pred) = 81.67 %	

Fuente	DF	SS	MS	F	P
Regresión	2	75.695	37.848	173.39	0.000
Error	25	5.457	0.218		
Total	27	81.152			
T	1	29.570			
ln(T)	1	46.125			

Figura 4.28: Diagnósticos de SIDA en el tiempo T. Género femenino



Variable de Respuesta: CH (Contacto Heterosexual)

Predictora: Tiempo

Transformación normal

La ecuación de regresión es:

$$\ln(CH) = - 1.01 + 0.776 T - 0.0213 T^2$$

Predictor	Coef	SE Coef	T	P
constante	-1.0077	0.3020	-3.34	0.003
T	0.77585	0.04800	16.16	0.000
T^2	-0.021308	0.001606	-13.26	0.000

S = 0.495132	R-Sq = 93.5 %	R-Sq(adj) = 93.0 %
PRESS = 7.76068	R-Sq(pred) = 91.82 %	

Fuente	DF	SS	MS	F	P
Regresión	2	88.697	44.349	180.90	0.000
Error	25	6.129	0.245		
Total	27	94.826			
T	1	45.563			
T^2	1	43.134			

Transformación gamma

Los resultados de la transformación gamma se presentan a continuación:

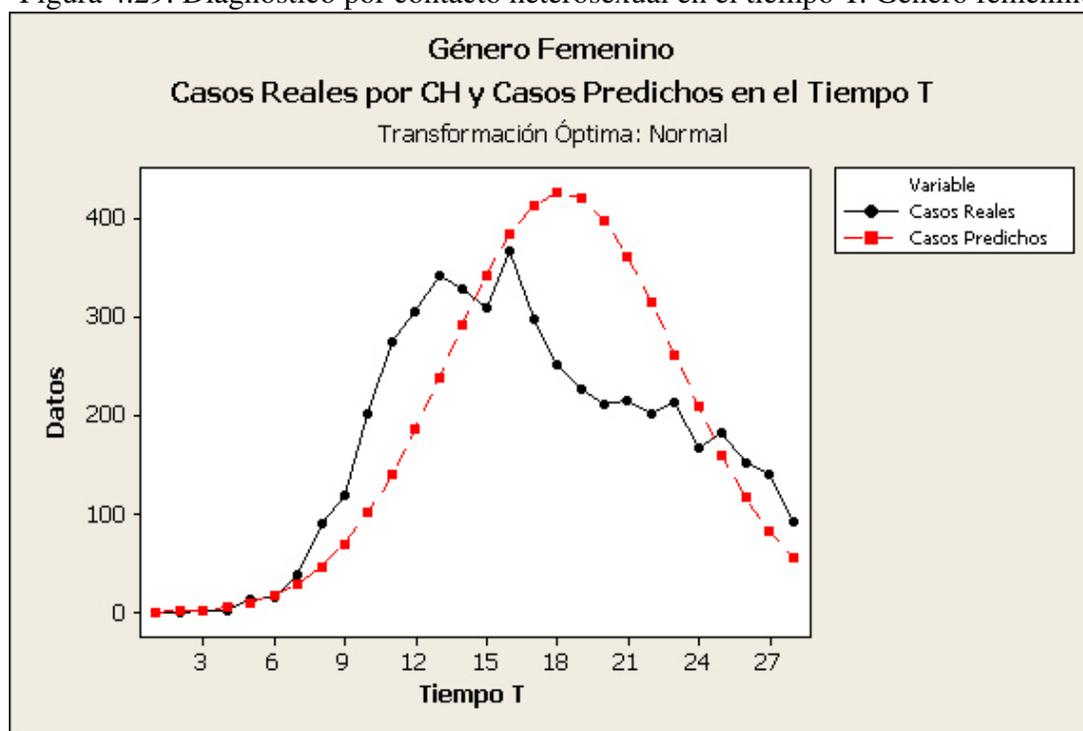
$$\ln(CH) = - 1.65 - 0.178 T + 3.55 \ln(T)$$

Predictor	Coef	SE Coef	T	P
constante	-1.6541	0.5278	-3.13	0.004
T	-0.17849	0.04368	-4.09	0.000
$\ln(T)$	3.5527	0.4255	8.35	0.000

S = 0.721150	R-Sq = 86.3 %	R-Sq(adj) = 85.2 %
PRESS = 32.1366	R-Sq(pred) = 66.11 %	

Fuente	DF	SS	MS	F	P
Regresión	2	81.824	40.912	78.67	0.000
Error	25	13.001	0.520		
Total	27	94.826			
T	1	45.563			
ln(T)	1	36.261			

Figura 4.29: Diagnóstico por contacto heterosexual en el tiempo T. Género femenino



Variable de Respuesta: UDI (Usuarios de Drogas Intravenosas)

Predictora: Tiempo

Los resultados de las transformaciones tanto a gamma como a normal se detallan a continuación:

Transformación normal

$$\ln(UDI) = 0.010 + 0.662 T - 0.0201 T^2$$

Predictor	Coef	SE Coef	T	P
constante	0.0100	0.3954	0.03	0.980
T	0.66177	0.06027	10.98	0.000
T^2	-0.020111	0.001957	-10.28	0.000

S = 0.550723	R-Sq = 83.8 %	R-Sq(adj) = 82.4 %
PRESS = 10.3195	R-Sq(pred) = 77.01 %	

Fuente	DF	SS	MS	F	P
Regresión	2	37.613	18.806	62.01	0.000
Error	24	7.279	0.303		
Total	26	44.892			
T	1	5.591			
T^2	1	32.022			

Transformación gamma

La ecuación de regresión obtenida fue la siguiente:

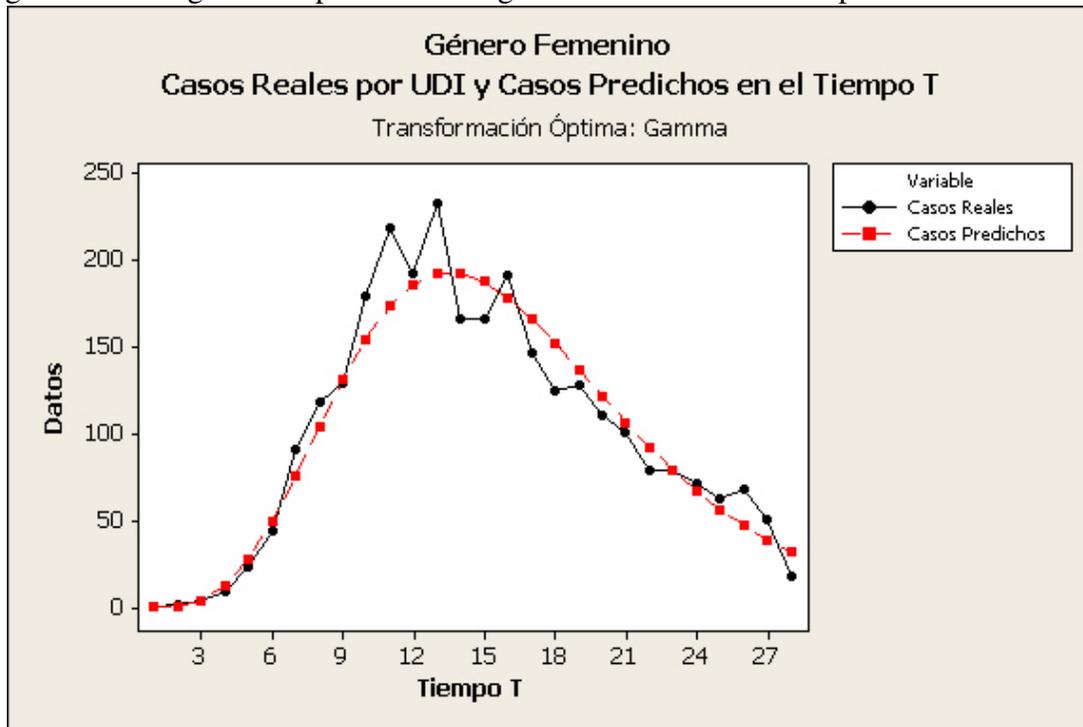
$$\ln(UDI) = - 3.27 - 0.392 T + 5.31 \ln(T)$$

Predictor	Coef	SE Coef	T	P
constante	-3.2694	0.2578	-12.68	0.000
T	-0.39198	0.01734	-22.61	0.000
$\ln(T)$	5.3055	0.1935	27.43	0.000

S = 0.225025	R-Sq = 97.3 %	R-Sq(adj) = 97.1 %
PRESS = 2.10414	R-Sq(pred) = 95.31 %	

Fuente	DF	SS	MS	F	P
Regresión	2	43.677	21.838	431.28	0.000
Error	24	1.215	0.051		
Total	26	44.892			
T	1	5.591			
$\ln(T)$	1	38.086			

Figura 4.30: Diagnósticos por uso de drogas intravenosas en el tiempo T. Género femenino



Variable de Respuesta: TRANS (Transfusiones)

Predictora: Tiempo

Transformación gamma

La ecuación de regresión obtenida es la siguiente:

$$\ln(TRANS) = -4.61 - 0.637 T + 5.74 \ln(T)$$

Predictor	Coef	SE Coef	T	P
constante	-4.613	3.225	-1.43	0.183
T	-0.6374	0.1908	-3.34	0.007
ln(T)	5.745	2.246	2.56	0.028

S = 0.525833	R-Sq = 72.9 %	R-Sq(adj) = 67.4 %
PRESS = 7.81183	R-Sq(pred) = 23.36 %	

Fuente	DF	SS	MS	F	P
Regresión	2	7.4281	3.7140	13.43	0.001
Error	10	2.7650	0.2765		
Total	12	10.1931			
T	1	5.6187			
ln(T)	1	1.8094			

La ecuación de regresión obtenida luego de efectuar la transformación a la distribución normal se muestra a continuación:

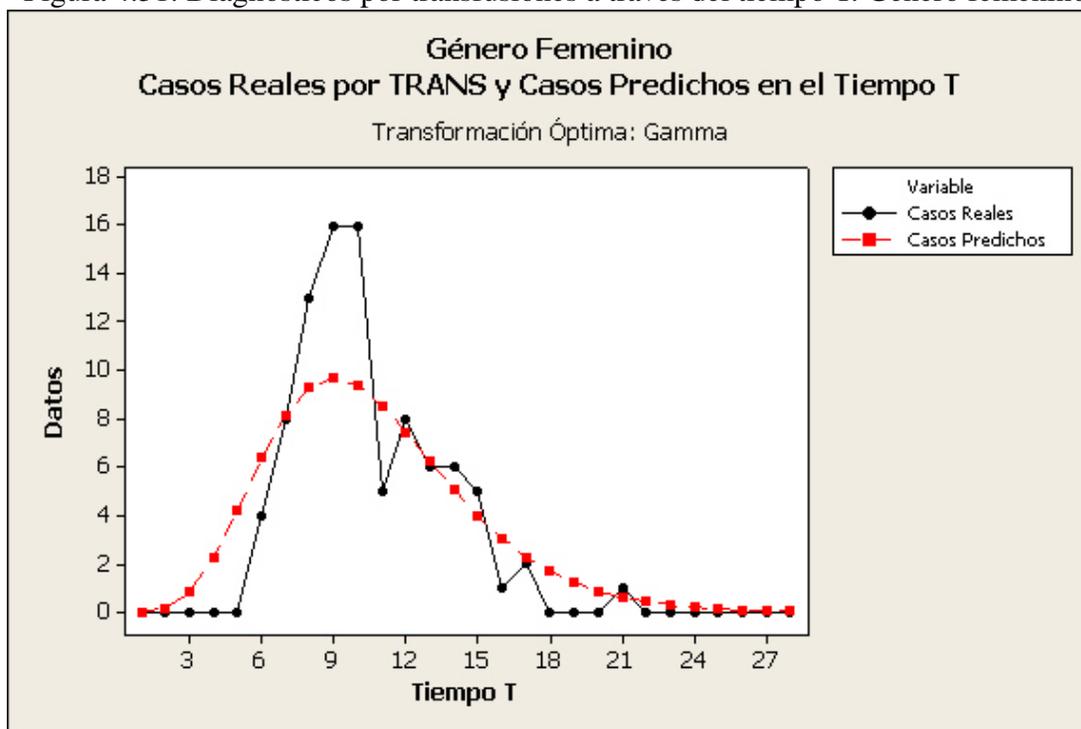
$$\ln(TRANS) = 1.23 + 0.242 T - 0.0153 T^2$$

Predictor	Coef	SE Coef	T	P
constante	1.231	1.359	0.91	0.386
T	0.2419	0.2206	1.10	0.299
T^2	-0.015343	0.008340	-1.84	0.096

S = 0.584611	R-Sq = 66.5 %	R-Sq(adj) = 59.8 %
PRESS = 13.7149	R-Sq(pred) = 0.00 %	

Fuente	DF	SS	MS	F	P
Regresión	2	6.7754	3.3877	9.91	0.004
Error	10	3.4177	0.3418		
Total	12	10.1931			
T	1	5.6187			
T^2	1	1.1567			

Figura 4.31: Diagnósticos por transfusiones a través del tiempo T. Género femenino



Variable de Respuesta: TV (Trasmisión Vertical)

Predictora: Tiempo

La ecuación de regresión obtenida para la transformación a una distribución normal es la siguiente:

$$\ln(TV) = 2.61 + 0.050 T - 0.00519 T^2$$

Predictor	Coef	SE Coef	T	P
constante	2.6106	0.8250	3.16	0.005
T	0.0503	0.1147	0.44	0.666
T^2	-0.005187	0.003563	-1.46	0.162

S = 0.655664	R-Sq = 61.3 %	R-Sq(adj) = 57.2 %
PRESS = 11.9885	R-Sq(pred) = 43.13 %	

Fuente	DF	SS	MS	F	P
Regresión	2	12.9122	6.4561	15.02	0.000
Error	19	8.1680	0.4299		
Total	21	21.0802			
T	1	12.0012			
T^2	1	0.9110			

La siguiente transformación aplicada fue la de distribución gamma, y los resultados se resumen a continuación:

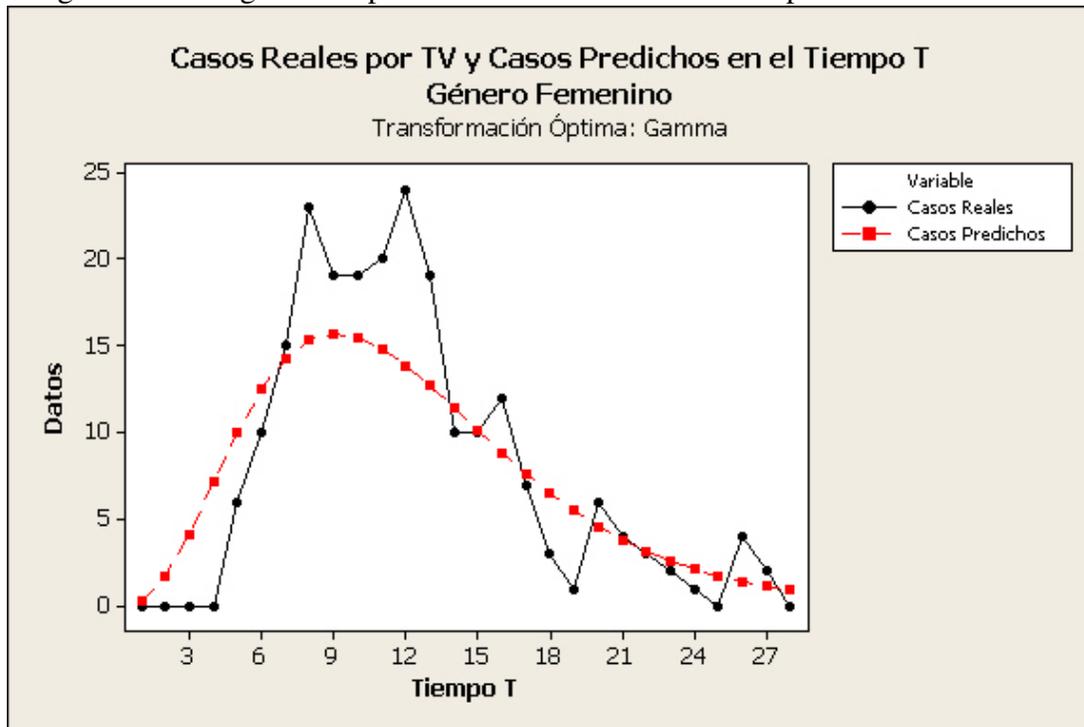
$$\ln(TV) = -0.89 - 0.330 T + 3.01 \ln(T)$$

Predictor	Coef	SE Coef	T	P
constante	-0.892	1.882	-0.47	0.641
T	-0.32976	0.08956	-3.68	0.002
$\ln(T)$	3.006	1.216	2.47	0.023

S = 0.601249	R-Sq = 67.4 %	R-Sq(adj) = 64.0 %
PRESS = 9.71326	R-Sq(pred) = 53.92 %	

Fuente	DF	SS	MS	F	P
Regresión	2	14.2117	7.1058	19.66	0.000
Error	19	6.8685	0.3615		
Total	21	21.0802			
T	1	12.0012			
ln(T)	1	2.2105			

Figura 4.32: Diagnósticos por trasmisión vertical en el tiempo T. Género femenino



ρ

Otras Variables

Otros tipos de contagio de SIDA en el género femenino no contenían mucha información, por lo cual no fue necesario hacer transformaciones para ver su tendencia. Esto se confirma además en el análisis de Stepwise para Mujeres así como en el análisis de mejores subconjuntos. Los diagrama de puntoss de estas variables se presentan a continuación:

Figura 4.33: Diagnósticos por hemofilia en el tiempo T. Género femenino

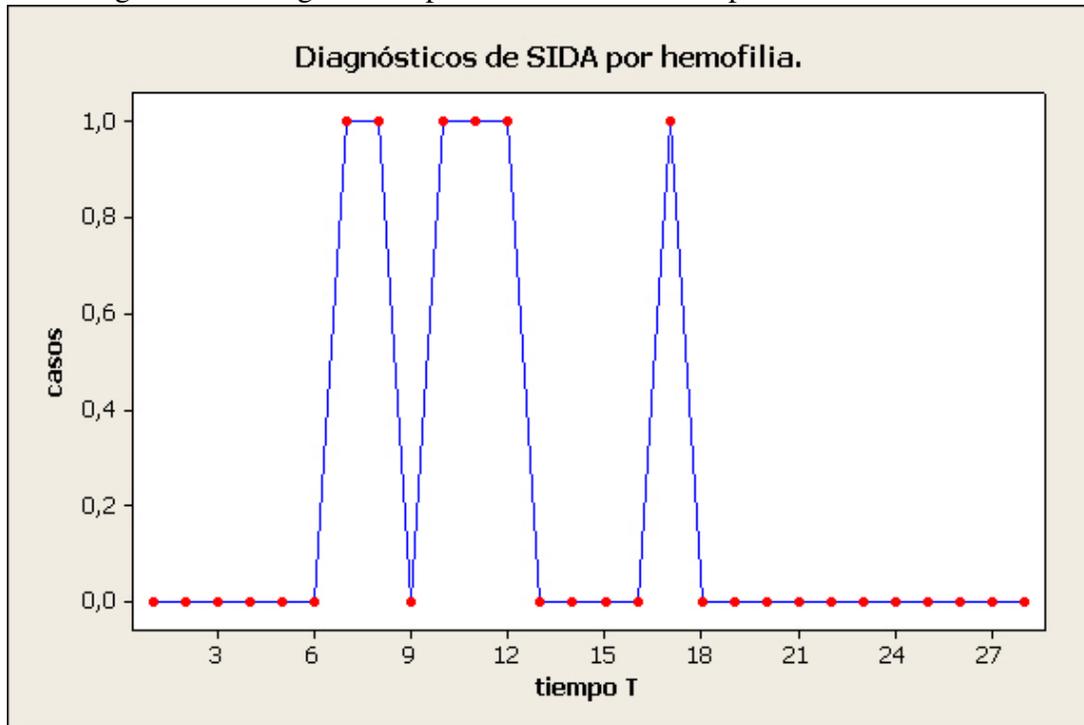


Figura 4.34: Diagnósticos por transfusiones pediátricas en el tiempo T. Género femenino

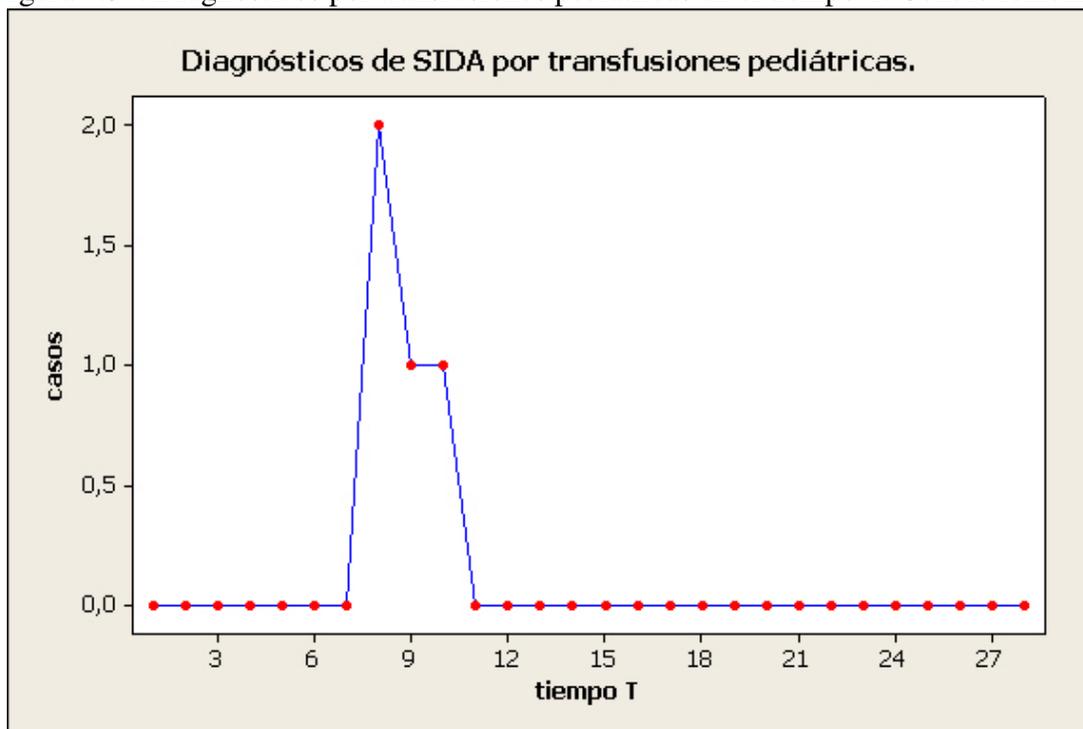


Figura 4.35: Diagnósticos por hemofilia pediátrica en el tiempo T. Género femenino

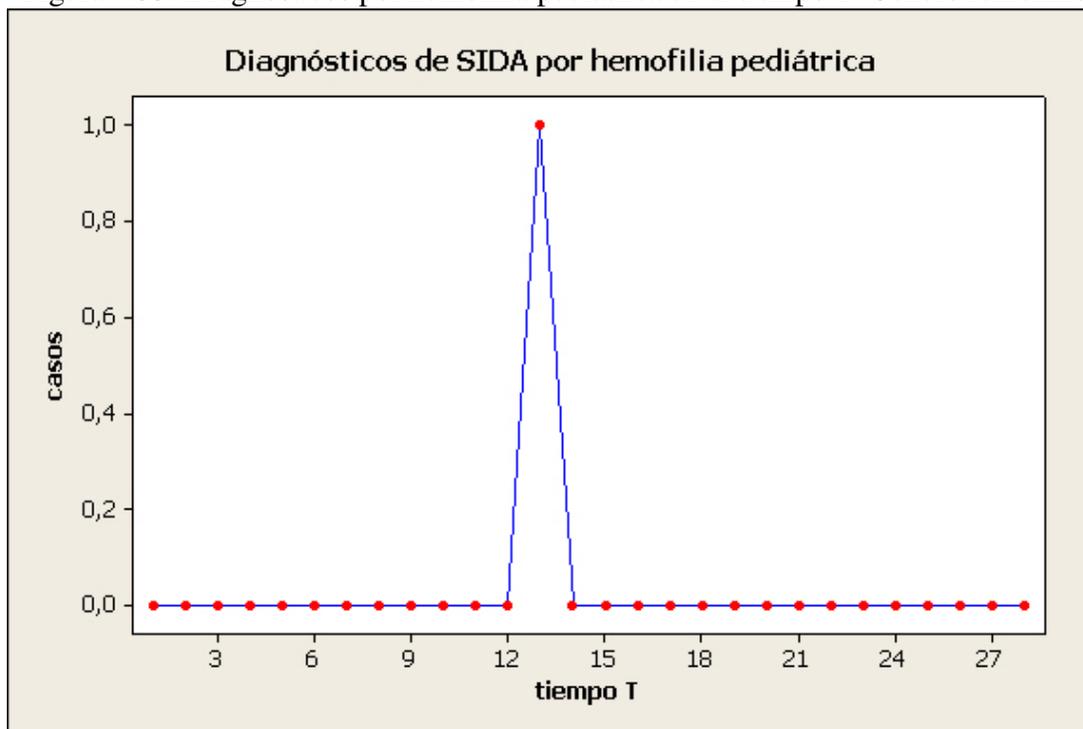
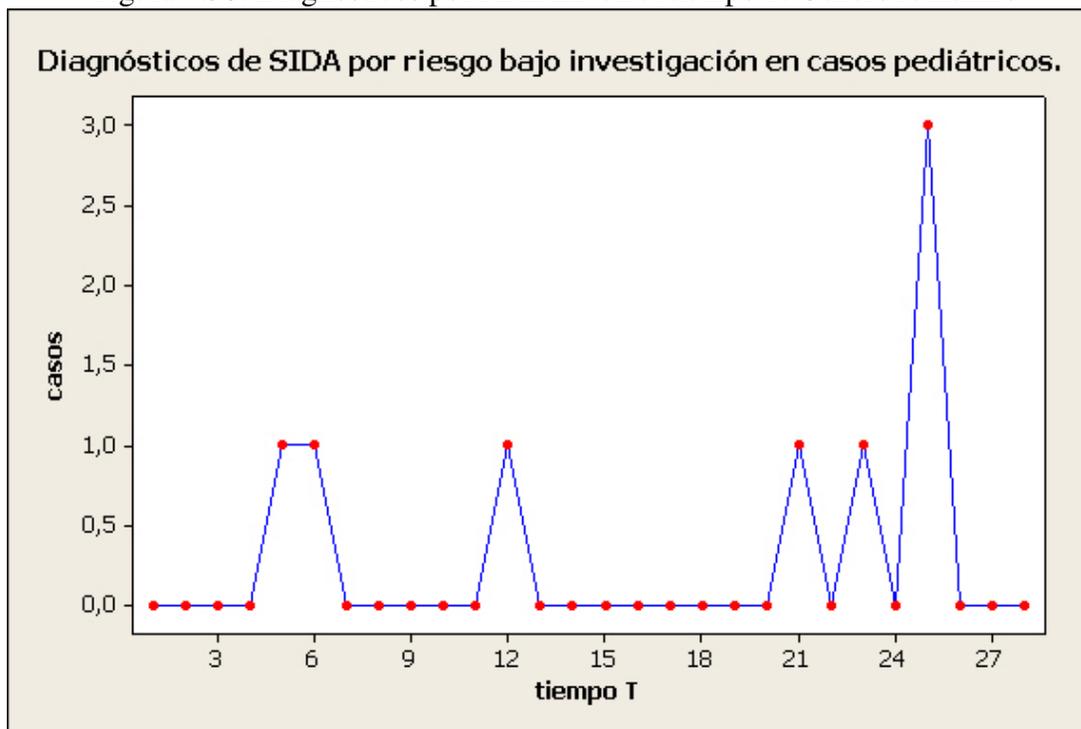


Figura 4.36: Diagnósticos por RBIPED en el tiempo T. Género femenino



4.8. Comparación de resultados de transformaciones

Tabla 4.17: Coeficientes de determinación para las transformaciones de variables predictoras

	Hombre normal	Hombre gamma	Mujer normal	Mujer gamma	TOTAL normal	TOTAL gamma
Muertes	84.2 %	96.1 %	88.1 %	94.6 %	85.3 %	96.9 %
Casos	82.3 %	98.6 %	90.1 %	93.3 %	84.9 %	97.8 %
HSH	84.2 %	97.2 %	SFD	SFD	84.2 %	97.2 %
UDI	82.6 %	96.2 %	83.8 %	97.3 %	83 %	96.6 %
HSHUDI	73.1 %	96.7 %	SFD	SFD	73.1 %	96.7 %
CH	81.5 %	94.3 %	93.5 %	86.3 %	94.2 %	86.1 %
HEM	81.4 %	86.1 %	SFD	SFD	80.7 %	86.8 %
TRANS	36,3 %	55 %	66.5 %	72.9 %	63 %	83.3 %
RBI	N/A	N/A	N/A	N/A	N/A	N/A
TV	57 %	74.7 %	61.3 %	67.4 %	65.1 %	84.1 %
TPED	N/A	N/A	N/A	N/A	N/A	N/A
HPED	N/A	N/A	N/A	N/A	N/A	N/A
RBIPED	N/A	N/A	N/A	N/A	N/A	N/A

En la tabla 4.17 se observan los coeficientes de determinación resultante de cada transformación aplicada a las variables predictoras, tanto para el Caso total como para los casos clasificados por género. Por ejemplo, para la variable predictora UDI, en el caso del género masculino, el mejor modelo para explicar las muertes de individuos de este género, en función del tiempo T, está dada por una ecuación de la forma de la distribución gamma, siendo su coeficiente de determinación 96.2 %.

CAPÍTULO 5

Conclusiones

1. El modelo lineal óptimo para predecir las muertes por SIDA está dado por $muertes = 19,6 + 1,26UDI$ para el caso total; $muertes = -10,4 + 1,01UDI + 0,89CH$ para el género masculino; y por $muertes = 9,6 + 0,96CH + 2,34TV - 2,28T$ para el género femenino.
2. Los resultados de los análisis de independencia indican que el tipo de contagio de SIDA y el género del individuo son variables dependientes, siendo el género masculino el más afectado por esta enfermedad, y siendo el uso de drogas intravenosas la principal forma de contagio de la misma. El tipo de contagio de SIDA y el grupo de edad al que pertenece el individuo también resultó que son variables dependientes, siendo más susceptibles a la misma los individuos que se encuentran entre 35 y 44 años de edad.
3. Con respecto al análisis comparativo de las transformaciones efectuadas en las predictoras, se concluye que las variables muertes, casos diagnosticados, HSH (sólo para el caso de género masculino y el caso global), usuarios de drogas intravenosas, los usuarios de drogas intravenosas y de contacto homosexual, hemofilia, y transmisión vertical, se

comportan como una variable gamma a través del tiempo.

4. Por otro lado, la variable contacto heterosexual (CH) para el caso de los hombres y para el caso total, tienen a comportarse como una gamma, sin embargo para el caso particular de mujeres, se comporta como una normal a través del tiempo.
5. Para el caso total, el análisis de regresión escalonada hacia adelante indica que la variable de mayor importancia fue UDI. Las variables CH, T, HSH, y HSHUDI fueron eliminadas por presentar problemas de multicolinealidad. El α de ingreso de variable al modelo fue fijado en 0.05
6. Para el género masculino fueron de importancia las variables UDI, y CH. Las variables HSH, HSHUDI, HEM y T fueron eliminadas por presentar problemas de multicolinealidad; mientras que para el género femenino, las variables relevantes de acuerdo al análisis de stepwise fueron CH, T y TV. La variable UDI fue eliminada por presentar problemas de multicolinealidad.
7. En el análisis de correlaciones se observa que las variables CH y UDI están altamente correlacionadas (0.792), al igual que CH y HSH (0.979). Por género la correlación más alta se da entre las variables CH y UDI (0.864) para el género femenino; y entre CH y UDI (0.736), y CH y HSH(0.842). Para evitar el problema de multicolinealidad en el modelo, se usó simultáneamente el análisis de regresión escalonada, y las correlaciones entre las variables predictoras, para determinar las variables que se incluyeron en el modelo.
8. Como en el modelo de regresión múltiple de muertes versus todos los tipos de infección y el tiempo, presentó problemas de multicolinealidad, autocorrelación, y varianza no constante, se descartó resultando únicamente en los modelos descritos en la conclusión 1.

Bibliografía

- [1] Acuna, E. Análisis de Regresión. (2007). <http://math.uprm.edu/edgar> Consultado en marzo del 2009.
- [2] Anderson, R. (1988). The Epidemiology of VIH Infection: Variable Incubation Plus Infectious Periods and Heterogeneity in Sexual Activity. *J.R. Statist. Soc. A.* 151, Part 1, pp 66-93
- [3] Anderson, R. Medley, G. (1988). Epidemiology of VIH infection and AIDS: incubation and infectious periods, survival and vertical transmission. *AIDS* 1988, 2 suppl 1): S57-S63
- [4] Bailey, N. (1988). Simplified Modelling of the Population Dynamics of VIH/AIDS.
- [5] Bailey, N. Estreicher, F. (1988). Simplified Modelling of the Population Dynamics of VIH/SIDA, Hopital Cantonal Universitaire, Geneva, Switzerland.
- [6] Baryarama, F. Mugisha, J. Luboobi, V. (2006). Mathematical Model for VIH/SIDA with complacency in a population with declining prevalence. *Computational and Mathematical Methods in Medicine*, vol 7, No 1.
- [7] Bautista, E. Thompson, E. Guzmán, C. Ramírez, D. Figueroa, J. (2006). Forecasting of Multiple Points in Time Series with Support Vector Machines.
- [8] Blattner W. (1991). VIH Epidemiology: Past, present and future. *Viral Epidemiology Section*, National Cancer Institute, Bethesda, Maryland.
- [9] Blower S. Hartel, D. Dowlatabadi, H. Anderson, R. May, R. (1991). Drugs, sex and HIV: a mathematical model for New York City.
- [10] Blower S. , Schwartz E. , Mills J. (2003) Forecasting the Future of VIH Epidemics: the Impact of Antiretroviral Therapies and Imperfect Vaccines. *AIDS Rev* 2003;5: 113-125.
- [11] Blower S, Dowlatabadi H. (1994). Sensitivity and Uncertainty Analysis of Complex Models of Disease Transmission: an VIH Model as an Example. *International Statistical Review* Vol. 62, No. 2, pp. 229-243

- [12] Cancel, L. Robles, R. Colón, H. Matos, T. Freeman, D. (1994). VIH AIDS Risk perception, VIH Risk Behaviors and VIH Seropositivity among injection drug users in Puerto Rico.
- [13] Casella, G. Berger, R. Statistical Inference. Second Edition. Duxbury Advanced Series.
- [14] Centers for Disease Control and Prevention.
<http://www.cdc.gov/VIH/topics/surveillance/basic.htm> Consultado en Abril del 2009.
- [15] Chesson H. , Pinkerton S. (2000) Sexually Transmitted Diseases and the Increased Risk for VIH Transmission: Implications for Cost Effectiveness Analyses of Sexually Transmitted Disease Prevention Interventions. JAIDS Journal of Acquired Immune Deficiency Syndromes. 24: 48-56.
- [16] Chin Long Chiang Statistical Methods of Analysis.
- [17] Cichocki, M. <http://SIDA.about.com/cs/SIDAFactsheets/a/VIH101.htm> . Consultado en enero del 2009.
- [18] Duffin P, Tullis R. (2002), Mathematical Models of the Complete Course of VIH Infection and SIDA. Journal of Theoretical Medicine, vol4, pp 215-221.
- [19] Garnett G. An introduction to mathematical models in sexually transmitted disease epidemiology.
- [20] Gómez M, Fernández D, Otero J, Miranda S. Hunter R. (2000). The shape of the VIH/SIDA epidemic in Puerto Rico. Rev Panama Salud Publica.
- [21] Góngora, R. (2006). VIH y SIDA: Un enfoque a 25 años de su emergencia. Grupo de Estudios de VIH/SIDA del Sureste, A.C. Merida, Yucatán, Mexico.
- [22] Isham V. (1988). Mathematical Modelling of the Transmission Dynamics of VIH Infection and SIDA: A Review. Journal of the Royal Statistical Society. Serie A. Vol 151, No 1, pp 5-49.
- [23] Karon J, Kaplan EH, Brookmeyer R, Song R. (2002). Proposals for estimating VIH incidence in the United States from VIH reporting data and the detuned assay. Int Conf AIDS, abstract no WeOrC1308.
- [24] Kingman J. Durban J. Cox D. Healy M. (1998). Statistical Requirements of the SIDA Epidemic. Journal of the Royal Statistical Society. Series A. vol 151, No 1, pp 127-130.
- [25] Lawrence F. (1999), Estimated HIV Risk Among Hispanics in a National Sample of Drug Users.

- [26] Leaman D., (2007) Severe Anxiety Symptomatology and HIV Risk Behavior Among Hispanic Injection Drug Users in Puerto Rico
- [27] Levin Bruce, Bull J, Stewart Frank. (2001) Epidemiology, Evolution, and Future of the VIH AIDS Pandemic. *Emerging Infectious Diseases*. Vol. 7, No. 3 Supplement, June 2001.
- [28] Lippincott. Williams. Wilkins. (2005). *Steadman's Medical Dictionary for the health professions and nursing*. 5th edition. Philadelphia.
- [29] Montesinos, M. Hernandez, C. (2007) Modelos matematicos para enfermedades infecciosas. *Salud Publica de Mexico*, vol 49, no.3.
- [30] Montoya, I. Bell, D. Richard, A. Carlson, J. Trevino, R. (1999). Estimated VIH Risk Among Hispanic in a National Sample of Drug Users.
- [31] Pengfei, L. (2005). *Box-Cox Transformations: An Overview*. Department of Statistics, University of Connecticut.
- [32] Perelson, A. Nelson, P. (1999). *Mathematical Analysis of VIH-1. Dynamics in Vivo*. *SIAM Review*. Vol 41, No 1, pp 3-44.
- [33] Departamento de Salud, Programa Vigilancia del SIDA. Estado Libre Asociado de Puerto Rico.
- [34] Quinn, T. , Overbaugh, J. (2005). VIH-AIDS in Women: An Expanding Epidemic. *SCIENCE*, vol 308, June 10th, 2005.
- [35] Reyes, J. Robles, R. Colom, H. Marrero, A. Matos, T. Calderon, J. Shepard, E. (2006). Severe Anxiety Symptomatology and VIH Risk Behavior Among Hispanic Injection Drug Users in Puerto Rico.
- [36] Robles, R. Marrer, C. Reyes, JC. Colón, H. Matos, T. Finlinson, H. (1998). Risk behaviors, VIH seropositivity and tuberculosis infection.
- [37] Robles, J. Robles, R. Colón, H. Matos, T. Finlinson, H. Marrero, C. (2005). Homelessness and VIH risk behaviors among drug injectors in Puerto Rico.
- [38] TendenciasPr. Reporte de Estadísticas Vitales. <http://www.tendenciaspr.com>.
- [39] Vera, Carmen. (1988) Comportamiento Riesgo Sexual, HIV, y Tuberculosis Infección.
- [40] Wilkie, A. (1988) An Actual Model for SIDA. *Journal of the Royal Statistical Society. Series A-* Vol 151, No , pp 35-99.