
# ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

## Instituto de Ciencias Matemáticas

## INGENIERÍA EN ESTADÍSTICA INFORMÁTICA

"ANÁLISIS DE VARIANZA FUNCIONAL PARA MEDIR EL EFECTO DE RECUBRIMIENTOS REVESTIBLES DE ALMIDÓN EN CARACTERÍSTICAS FÍSICO-QUÍMICAS DE PAPAYAS DE LA ESPECIE CARICA PAPAYA L., DURANTE LAS DOS PRIMERAS SEMANAS DEL PERÍODO DE MADURACIÓN POST-COSECHA"

## TESIS DE GRADO

Previa a la obtención del Título de:

## INGENIERO EN ESTADÍSTICA INFORMÁTICA

Presentado por:

Christian Eduardo Galarza Morales

Guayaquil – Ecuador

2012

# AGRADECIMIENTO

A mis dos grandes maestros y padres en la Estadística: el M.Sc. Gaudencio Zurita y el Ph.D. Francisco Vera.

# DEDICATORIA

A mis dos grandes amores:

*Esther y Silvia.*

# TRIBUNAL DE GRADUACIÓN

M.Sc. Vanessa Salazar Villalva
**PRESIDENTE DEL TRIBUNAL**

Ph.D. Francisco Vera Alcívar
**DIRECTOR DE TESIS**

Ing. Félix Ramírez Cruz
**VOCAL PRINCIPAL**

# DECLARACIÓN EXPRESA

“La responsabilidad del contenido de esta Tesis de Grado, me corresponde exclusivamente; y el patrimonio intelectual de la misma a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL’’.

(Reglamento de Graduación de la ESPOL).

_____

Christian Eduardo Galarza Morales

# RESUMEN

El mercado actual demanda productos de alta calidad y de consumo inmediato, que contengan sólo ingredientes naturales. El éxito de estos productos se debe a sus buenas características sensoriales y nutricionales. Por ello, ha sido necesario el desarrollo de nuevas tecnologías de conservación que retrasen los procesos de deterioro y mantengan la calidad e inocuidad microbiológica, durante la comercialización de estos productos.

Una alternativa es el uso de películas comestibles biodegradables las cuales crean una barrera a los gases, produciendo una atmósfera modificada alrededor del producto reduciendo la tasa de respiración y la pérdida de agua, aumentando así, la vida de “anaquel”.

La orientación de este estudio es realizar un Análisis de Varianza Funcional con el objetivo de determinar la efectividad del uso de dos diferentes tipos de películas comestibles en frutas al preservar el producto durante las dos primeras semanas de su período de maduración pos-cosecha. La ventaja de la aplicación de un Análisis de Datos Funcionales (FDA), es explotar las características que tienen las funciones, como poder realizar todas las operaciones matemáticas básicas entre ellas, el poder aplicar toda técnica

estadística univariada como multivariada, así como utilizar las propiedades de sus funciones derivadas y diferenciales, las cuales son una fuente de información adicional de los aspectos dinámicos del modelo. El lector deberá tener conocimientos previos en Álgebra Lineal, Estadística Multivariada, Simulación Matemática y Diseño de Experimentos.

# ÍNDICE GENERAL

# ÍNDICE DE GRÁFICOS

# ÍNDICE DE TABLAS

# INTRODUCCIÓN

La papaya es una fruta de alto perecimiento, por lo que el control de la maduración es esencial para aumentar la vida útil, tanto para el mercado interno como para la exportación [2]. El estado de madurez de la fruta al ser cosechados, es especialmente importante para su manejo, transportación y comercialización ya que repercute directamente en su calidad y potencial de almacenamiento [7].

Comúnmente conocidas como películas comestibles, estas consisten en un revestimiento o envoltorio producido a partir de almidón de maíz [2] donde éstas se utilizan como una cubierta sobre los frutos en el momento que éstos son cosechados.

El mecanismo por el cual estas películas conservan la calidad de frutas y vegetales es debido a que crean una barrera ante los gases, produciendo una atmósfera modificada alrededor del producto. Esta atmósfera reduce la disponibilidad de O e incrementa la concentración de $CO_2$. De tal forma, se reduce la tasa de respiración y la pérdida de agua, aumentando así, la vida de “anaquel”.

En el laboratorio de Bromatología de la Facultad de Ingeniería Mecánica y Ciencias de la Producción de la ESPOL, se llevó a cabo un diseño experimental con el fin de determinar la proporción adecuada de aceites esenciales (clavo de olor y canela) a aplicarse en películas comestibles, con el fin de maximizar la protección anti-fúngica del recubrimiento, creando una película comestible con aceites esenciales.

Si bien esta película maximiza la protección anti hongos de la papaya, la aplicación de estos aceites esenciales podría tener efectos desconocidos o tal vez adversos a las características físicas y químicas de la papaya, las cuales son percibidas por el cliente en la calidad del producto como la textura, color, peso, nivel de acidez pH y sólidos solubles en el producto.

La orientación de este estudio es utilizar análisis de varianza funcional con el objetivo de determinar la efectividad del uso de diferentes tipos de películas

comestibles en frutas al preservar el producto durante sus dos primeras semanas del período de maduración pos-cosecha. El uso de esta técnica funcional permitirá estimar los efectos para todo instante de tiempo $t$ durante el período de maduración, aún en los momentos donde no se realizaron mediciones.

Todos los gráficos, cálculos y programación fueron realizados en el software libre R, haciendo uso de la librería `fda` que contiene las técnicas estadísticas del Análisis de Datos Funcional.

# CAPÍTULO I

## 1. MOTIVACIÓN

### *1.1. Introducción*

En FDA los datos discretos son transformados en funciones a través de un proceso de interpolación utilizando bases ortogonales de funciones como polinomios, *splines*, exponenciales, Fourier, entre otras, dado que en la práctica los datos son tomados como muestras discretas; ya sea en un estudio longitudinal donde la continuidad corresponde al tiempo $t$ donde se toman las mediciones, o un estudio geoespacial donde la continuidad está relacionada a la posición espacial de las observaciones [8].

### *1.2. Características*

La aplicación de análisis de datos funcionales tiene características deseables que no se tendrían si al contrario realizáramos un análisis estadístico de datos individuales. Algunas de las principales características mencionadas por [6] son:

- Los datos constituyen un conjunto de mediciones continuas.
- Se puede suponer que el proceso que genera los datos es suave y continuo.
- No se requieren supuestos paramétricos para la modelación.
- A menudo se tiene múltiples medidas de un mismo proceso.
- Por lo general, los procesos tienen alta afijación y poco ruido.
- Las funciones son diferenciables y derivables por lo que sus funciones derivadas son una fuente adicional de información acerca de los aspectos dinámicos del modelo.
- Cada dato funcional o curva constituye una observación, al contrario de que los puntos individuales que forman la curva sean considerados como observaciones.
- Toda técnica estadística, ya sea descriptiva, inferencial o multivariada puede ser aplicada para datos funcionales.

## *1.3. Estadísticas descriptivas para datos funcionales*

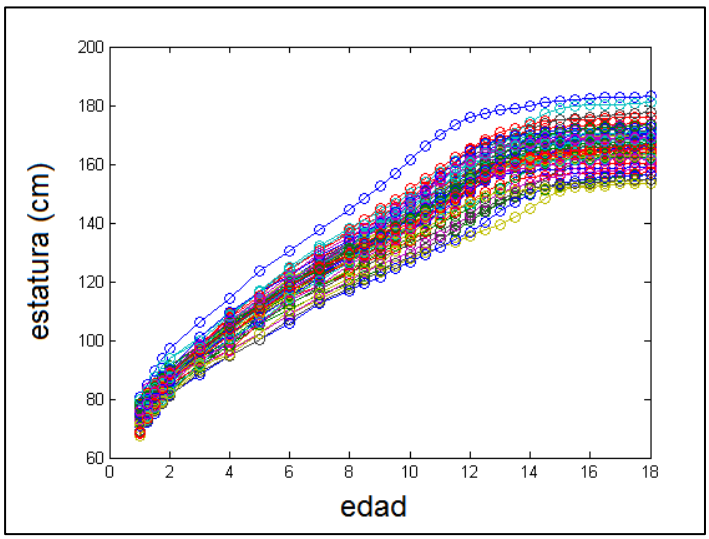
Las estadísticas descriptivas para datos univariados que resulta familiar en las clases introductorias de estadística son aplicadas de igual manera para datos funcionales. El gráfico 1.1 muestra las estaturas de 20 niñas medidas en 31 ocasiones desde su nacimiento

hasta los 18 años de edad, esto para el *Estudio de Crecimiento de la Universidad de Berkeley*.

Las mediciones no se encuentran igualmente espaciadas; durante el primer año de vida las mediciones fueren trimestrales, medidas anuales desde los dos hasta los ocho años y posteriormente las mediciones fueron de manera semestral.

Los círculos indican la medición observada. Sin problema alguno se podría hablar de funciones de crecimiento las cuales se calcularon interpolando las mediciones observadas por medio de funciones bases.

**Gráfico 1.1:** Funciones de crecimiento resultantes de interpolar las estaturas de 20 niñas medidas en 31 ocasiones desde su nacimiento hasta los 18 años de edad.




**Fuente:** [5] Hooker, Giles: Introduction to Functional Data Analysis, International Workshop on Statistical Modeling, Cornell University.

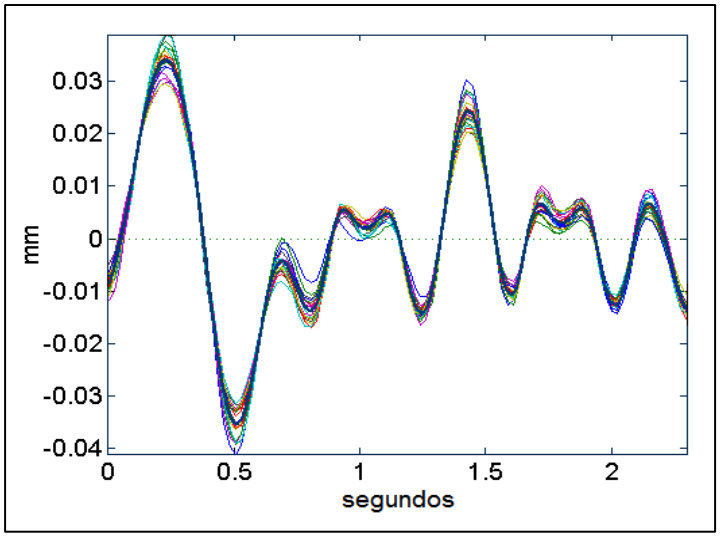
### *1.3.1. Media funcional*

La función media puede ser calculada como

$$\bar{x}(t) = \frac{1}{n}\sum x_i(t), \qquad (1.1)$$

donde $\bar{x}(t)$ es el promedio de las $n$ curvas correspondientes a las $n$ observaciones.

**Gráfico 1.2:** Funciones correspondientes al desplazamiento de una pieza mecánica de su eje en un ciclo completo de duración de 2.3 segundos. La curva sólida azul representa la función promedio del desplazamiento.




**Fuente:** [5] Hooker, Giles: Introduction to Functional Data Analysis, International Workshop on Statistical Modeling, Cornell University.

El gráfico 1.2 muestra la función media correspondiente al desplazamiento de una pieza mecánica de su eje en un ciclo completo de duración de 2.3 segundos.

### *1.3.2. Varianza funcional*

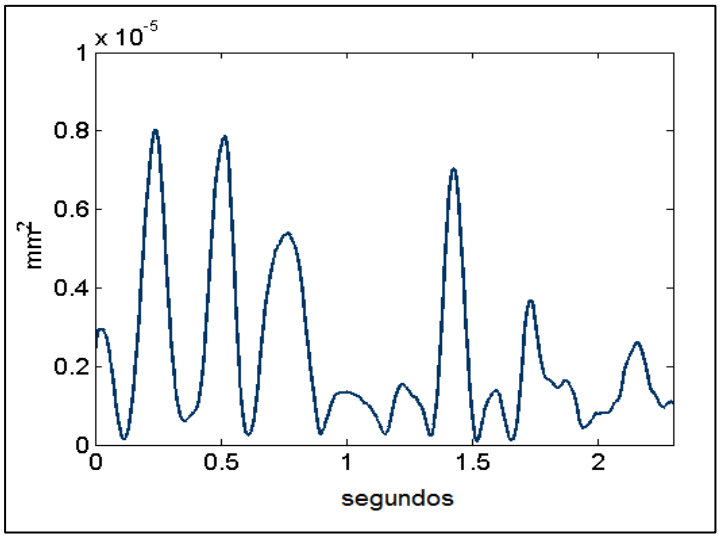
De igual manera la función varianza puede ser calculada como

$$s^2(t) = \frac{1}{n-1}\sum\left(x_i(t) - \bar{x}(t)\right)^2, \qquad (1.2)$$

donde la raíz cuadrada de esta función sería la función desviación estándar que explicaría la variación absoluta promedio de las curvas con respecto a la función media.

En el gráfico 1.3 se observa la función varianza correspondiente al desplazamiento de una pieza mecánica de su eje durante un ciclo completo de 2.3 segundos.

**Gráfico 1.3:** Función varianza correspondiente al desplazamiento de una pieza mecánica de su eje en un ciclo completo de duración de 2.3 segundos.




**Fuente:** [5] Hooker, Giles: Introduction to Functional Data Analysis, International Workshop on Statistical Modeling, Cornell University.

### *1.3.3. Función de covarianza*

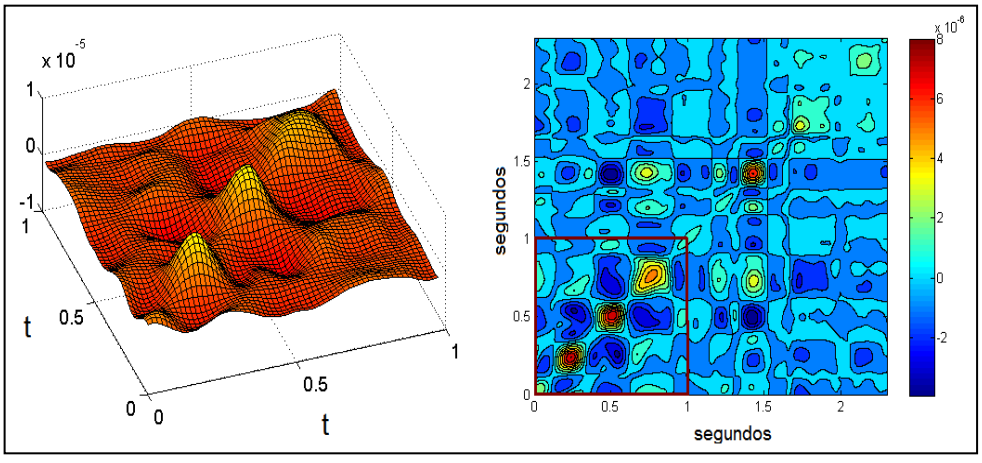
La función de covarianza mide la dependencia entre dos variables para toda unidad de tiempo $t$. Si la variable es la misma, se denomina funciónde autocovarianza, la cual mide la dependencia de la variable en diferentes instantes de tiempo $t_1$ y $t_2$.

La función de covarianza se obtiene como

$$C(t_1, t_2) = \frac{1}{n-1}\sum\big(x_i(t_1) - \bar{x}(t_1)\big)\big(y_i(t_2) - \bar{y}(t_2)\big). \qquad (1.3)$$

El gráfico 1.4 corresponde a la superficie de autocovarianza y diagrama de contornos para el ejercicio anterior.

**Gráfico 1.4:** Superficie de autocovarianza (izquierda) y diagrama de contornos (derecha) correspondiente al desplazamiento de una pieza mecánica de su eje durante un ciclo corto de 1 segundo y 2.3 segundos correspondientemente.




**Fuente:** [5] Hooker, Giles: Introduction to Functional Data Analysis, International Workshop on Statistical Modeling, Cornell University.

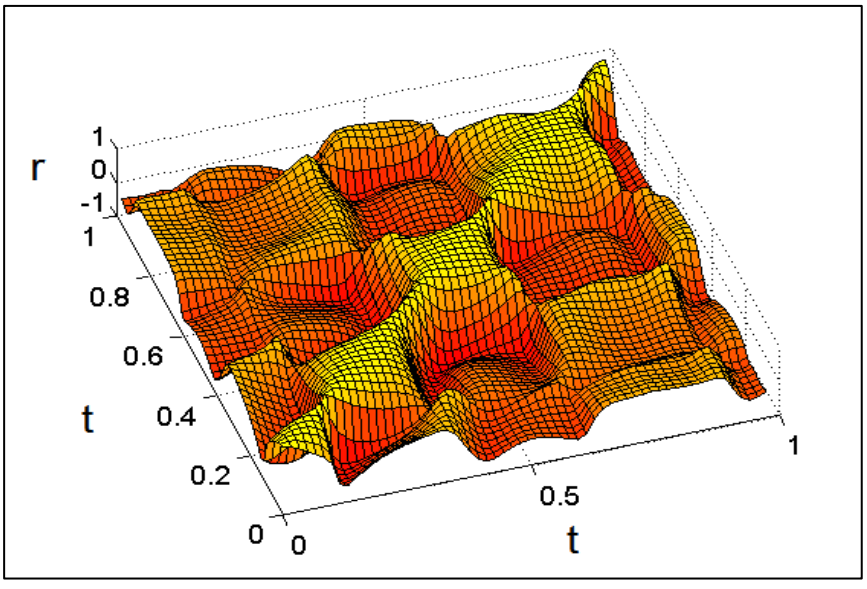
### *1.3.4. Función de correlación*

La función correlación se obtiene estandarizando la covarianza de igual manera que al trabajar con datos multivariados. Si se analiza la correlación entre una misma variable en diferentes instantes de tiempo, esta se llama función de autocorrelación.

En el gráfico 1.5, es relevante notar en la superficie la diagonal de unos (1's) cuando $t_1 = t_2$. La función correlación es calculada como

$$R(t_1, t_2) = \frac{C(t_1, t_2)}{\sqrt{s^2(t_1)s^2(t_2)}}\,. \qquad (1.4)$$

**Gráfico 1.5:** Superficie de autocorrelación correspondiente al desplazamiento de una pieza mecánica de su eje durante un ciclo de rotación de un segundo.




**Fuente:** [5] Hooker, Giles: Introduction to Functional Data Analysis, International Workshop on Statistical Modeling, Cornell University.

### *1.3.5. Prueba T funcional para la diferencia de medias de dos poblaciones independientes*

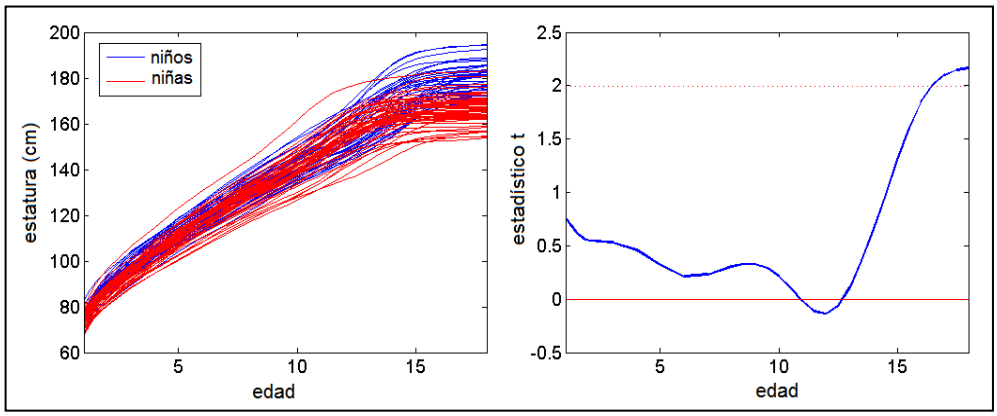
La prueba $T$ para diferencia de medias para el caso funcional potencia su nivel explicativo de tal manera que podremos determinar en qué instante de tiempo $t$ se podría considerar significativa la diferencia de medias y por consiguiente rechazar $H_0$.

El estadístico de prueba $T$ funcional bajo el supuesto de igualdad de varianzas se calcula como

$$T(t) = \frac{\bar{x}_1(t) - \bar{x}_2(t)}{s_p(t)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \qquad (1.5)$$

donde $s_p^2(t)$ es la función correspondiente al estimador $S^2$ "pooled".

**Gráfico 1.6:** En el panel izquierdo se muestran las 20 curvas de crecimiento representadas en azul las curvas correspondientes a los hombres y en rojo las mujeres. El panel derecho muestra la función T análoga al estadístico de prueba T.



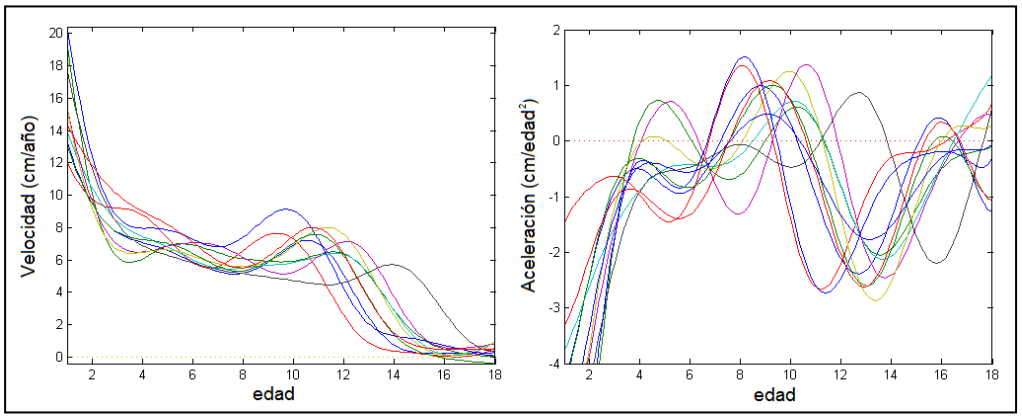

**Fuente:** [5] Hooker, Giles: Introduction to Functional Data Analysis, International Workshop on Statistical Modeling, Cornell University.

### *1.3.6. Derivadas como indicadores de la dinámica del modelo*

Si $x(t)$ es la estatura de un individuo en un instante de tiempo $t$ donde a la función $x(t)$ la llamaremos función estatura, haciendo uso de las propiedades de derivabilidad de las funciones podemos determinar la función velocidad de crecimiento como la primera derivada de $x(t)$ y la función aceleración como la segunda derivada de $x(t)$, esto es

$$v(t) = \frac{d}{dt}x(t), \qquad (1.6) \qquad\qquad acc(t) = \frac{d^2}{d^2t}x(t). \qquad (1.7)$$

**Gráfico 1.7:** En el panel izquierdo se muestran las 10 curvas correspondientes a la velocidad de crecimiento de las niñas mientras que en el panel derecho se encuentran las curvas de aceleración del crecimiento.




**Fuente:** [5] Hooker, Giles: Introduction to Functional Data Analysis, International Workshop on Statistical Modeling, Cornell University.

Es fácil notar los cambios en la velocidad y aceleración del crecimiento en las niñas especialmente cuando entran en la etapa de la pubertad y posteriormente cuando alcanzan la madurez.

### *1.3.7. Funciones lineales y regresión*

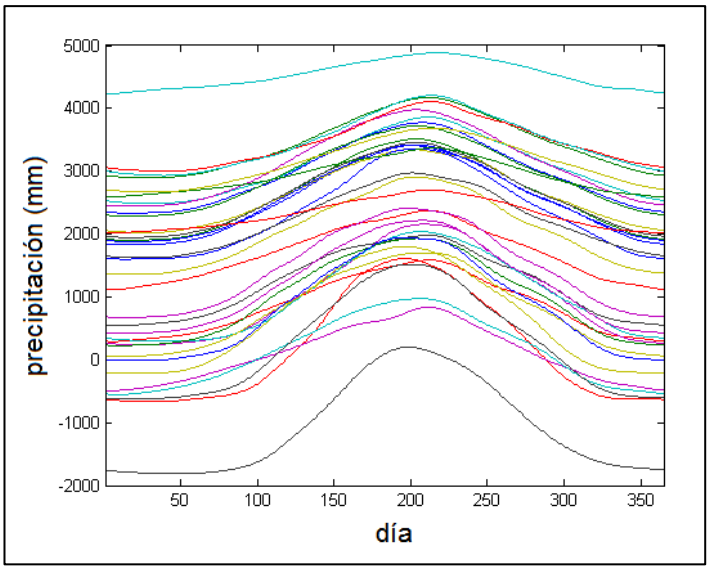
En regresión lineal estándar, el modelo estándar usualmente conocido se lo puede expresar como

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \epsilon_i,$$

mientras que en FDA, $x(t)$ es una variable de explicación para cada instante de tiempo $t$ donde

$$y_i(t) = \beta_0 + \int \beta(t) x_i(t) dt + \epsilon_i. \qquad (1.8)$$

**Gráfico 1.8:** Precipitación diaria promedio de lluvia en milímetros sobre Canadá. Las curvas fueron pronosticadas por medio de un modelo de regresión donde el estado, la temperatura y la estación climática fueron tomadas como variables de explicación.




**Fuente:** [5] Hooker, Giles: Introduction to Functional Data Analysis, International Workshop on Statistical Modeling, Cornell University.

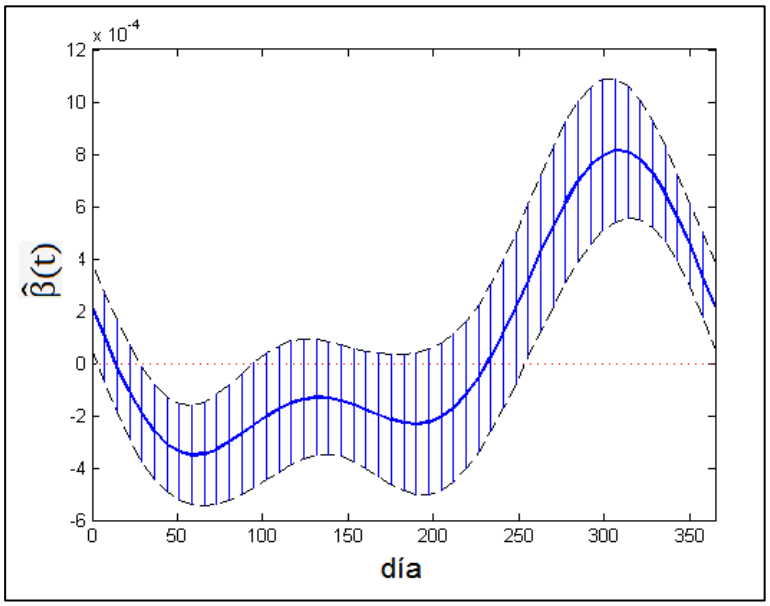
### *1.3.8. Intervalos de confianza para los estimadores de regresión*

Siempre es de vital interés medir el efecto de alguna de las variables de explicación sobre la variable de interés, mejor aún si podemos medir el comportamiento del efecto a lo largo del tiempo.

La característica funcional permite estimar el efecto aún para períodos donde no existen mediciones. Un intervalo de confianza funcional se calcula como:

$$b_i(t) - t_{\alpha/2} S_{b_i}(t) \leq \beta_i(t) \leq b_i(t) + t_{\alpha/2} S_{b_i}(t) \qquad (1.9)$$

**Gráfica 1.9:** Intervalo del 95% de confianza para la función $\beta(t)$ correspondiente al efecto de la temperatura sobre la log precipitación anual de lluvia en Canadá.




**Fuente:** [5] Hooker, Giles: Introduction to Functional Data Analysis, International Workshop on Statistical Modeling, Cornell University.

Al observar el gráfico 1.9, los valores negativos que toma el efecto para los dos primeros trimestres del año implica una disminución en las precipitaciones mientras que el valor positivo para el último trimestre corresponde a un aumento en las precipitaciones.

## *1.4. Generalidades*

Muchas de las técnicas estadísticas multivariadas se extienden al análisis de funciones como observaciones tales como análisis de componentes principales, análisis de conglomerados, correlación canónica, análisis discriminante, modelos lineales en general entre otras [6].

Por lo general, un estimador o estadístico resultante de la aplicación de alguna técnica estadística resulta ser una curva y por lo tanto nos permite hacer un análisis con respecto al tiempo o al espacio de su variación para todo instante de tiempo $t$ durante el período de análisis.

El análisis de datos funcionales hereda muchos de sus conceptos del análisis funcional, una rama del análisis matemático donde una función representa únicamente un punto en un espacio funcional [8]. Las

funciones son matemáticamente flexibles ya que con ellas se pueden realizar una amplia variedad de operaciones matemáticas. Además todas estas técnicas estadísticas funcionales se encuentran implementadas en softwares estadísticos y matemáticos como Matlab, R, SYSTAT, entre otros.

Se puede hacer referencia a diversas áreas de investigación donde los datos funcionales pueden ser encontrados, y las herramientas aquí mencionadas podrían potencialmente revelar relaciones que no sería fáciles de identificar con los métodos estadísticos usuales.

En algunos casos, el análisis de datos funcionales permitirá al investigador contestar preguntas no muy fáciles de responder, o que serían computacionalmente complicadas de responder usando los métodos estadísticos tradicionales [6]. Más aún, los modelos funcionales facilitan la representación visual de los datos, y esto provee un mayor poder de explicación.

# CAPÍTULO II

## 2. ANÁLISIS DE DATOS FUNCIONALES

### *2.1 Introducción*

El análisis de datos funcionales es una rama de la estadística que analiza los datos provenientes de curvas, superficies, entre otros, cuyo dominio sea continuo. “El dominio continuo” a menudo es el tiempo, pero también puede ser la ubicación espacial, pesos, frecuencias, probabilidades, etc.

La filosofía básica del análisis de datos funcionales es tomar las funciones observadas como entidades únicas mas no como una secuencia de observaciones individuales. El término funcional al referirse a observaciones hace referencia a la estructura intrínseca de los datos, mas no a su forma explícita [11].

En la práctica, los datos funcionales son $n$ pares ordenados $(t_j, y_j)$ observados de manera discreta donde $y_j$ es un valor de la función $x(t_j)$ en el tiempo $t_j$, usualmente afectado por un error $\epsilon_j$. Esto es expresado por

$$y_i = x(t_j) + \epsilon_j \tag{2.1}$$

o simplemente se puede expresar a (2.1) de manera algebraica como

$$\mathbf{y} = x(\mathbf{t}) + \mathbf{e}, \tag{2.2}$$

donde $\mathbf{y}$, $x(\mathbf{t})$ y $\mathbf{e}$ son vectores de longitud $n$.

Al hablar de datos funcionales no significa que exista un valor de $x$ para cada valor de $t$ pues esto implicaría una serie de infinitos valores, al contrario, se supone la existencia de una función $x(t)$ la cual aproxima los datos observados [11].

Además, usualmente deseamos definir que la función que "sigue" los datos es suave, de tal manera que un par valores adyacentes $y_j$ y $y_{j+1}$ se encuentren necesariamente relacionados entre sí, esperando no difieran en gran cantidad el uno del otro.

De no cumplirse el supuesto de suavidad, no existiría ventaja alguna entre realizar un análisis funcional o simplemente un análisis mutivariado. Sin embargo las curvas observadas no pueden ser del todo suaves dada la presencia de lo que usualmente se llama ruido o error de medición, donde se presenta sin patrón alguno y de manera impredecible por lo que por razones prácticas es ignorado, esto según Ramsay [11].

## 2.2 Representación de funciones por medio de sistemas bases

Un sistema de función base es un conjunto de funciones conocidas $\phi_k$ que son linealmente independientes entre sí y tienen la propiedad que pueden aproximar arbitrariamente sin problema alguno una función por medio de una combinación lineal de un número suficiente de $K$ funciones de estas [11]. Un sistema de función base representa una función x por una expansión lineal

$$x(t) = \sum_{k=1}^{K} c_k \phi_k(t). \tag{2.3}$$

### 2.2.1 Sistemas bases

La familia de sistemas de función base tradicional es el conjunto de monomios que son usualmente construidas por diversos tipos de bases. Las más comunes son las series de potencia

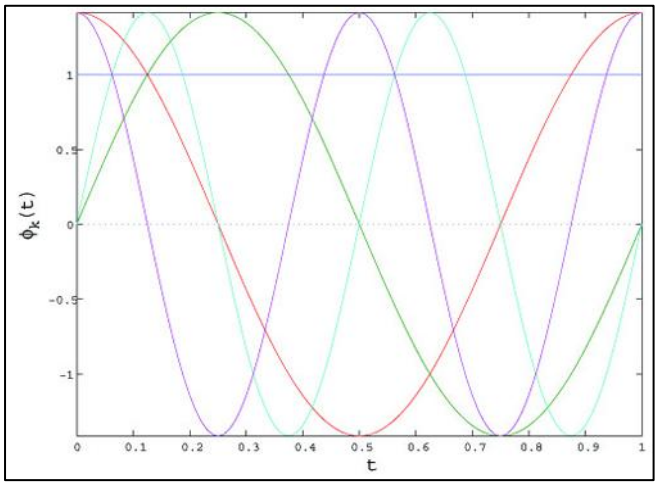
$$1, t, t^2, t^3, \dots, t^k, \dots \tag{2.4}$$

y junto a las series de potencia se encuentran los conocidos sistemas de series de Fourier,

$$1, sin(\omega t), cos(\omega t), sin(2\omega t), cos(2\omega t), sin(3\omega t), cos(3\omega t), \dots,$$
$$sin(k\omega t), cos(k\omega t), \dots \tag{2.5}$$

los cuales son de gran aplicabilidad, en casos donde los datos son de carácter periódico.

**Gráfico 2.1:** Cinco primeras funciones bases de una serie de Fourier con un periodo de una unidad.




**Fuente:** www.functionaldata.org [6]

Las funciones polinómicas son por tradición una de las más usadas; representada por

$$\phi_k(t) = (t - \omega)^k, k = 0, \dots, K \qquad (2.6)$$

donde $\omega$ es un parámetro de desplazamiento que usualmente es el punto medio del intervalo de estimación.

Existen otros tipos de bases como son las ondas (*wavelets*), polinómicas, poligonales, funciones escalón, constantes, empíricas, exponenciales, b-splines, entre otros, donde de estos dos últimos

se hablará con más detalle en el presente capítulo dado su aplicación en esta tesis.

#### *2.2.1.1 Sistemas bases exponenciales*

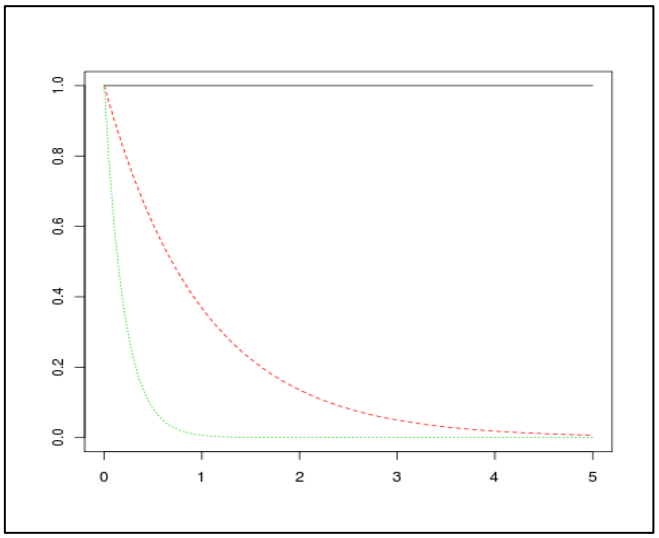
Los sistemas bases exponenciales consisten en una serie de funciones exponenciales,

$$e^{\lambda_1 t}, e^{\lambda_2 t}, \dots, e^{\lambda_k t}, \dots \tag{2.7}$$

donde los parámetros $\lambda_k$ son todos distintos, y usualmente $\lambda_1 = 0$ con el fin de representar una constante [11]. Cuando las funciones no suelen ser constantes, lineales, o simplemente no ondean, usualmente esta clase de comportamiento puede ser representado de manera acertada por las bases exponenciales cuyo comportamiento es seguido por muchos procesos naturales.

Sus aplicaciones son varias y algunas de ellas se fundamentan en que las ecuaciones diferenciales lineales con coeficientes constantes, tienen como soluciones expansiones en términos de funciones bases exponenciales, donde la diferenciabilidad de una función es una medida de suavidad.

**Gráfico 2.2:** Funciones bases exponenciales para el intervalo [0,5] con funciones bases 1, exp(-t) y exp(-5t).




### *2.2.1.2 Sistemas bases B-Splines*

Tal vez el sistema base que es utilizado con mayor frecuencia son los “B-splines” desarrollados por de Boor [3], dada las amplias ventajas que tiene sobre los sistemas clásicos. Alguna de ellas son:

- *Flexibilidad:* Las curvas usualmente tienen como características saltos, aspereza y comportamiento inesperado por lo que usando polinomios podrían seguir estos patrones solo con el uso de un gran número de funciones bases pero los splines no.

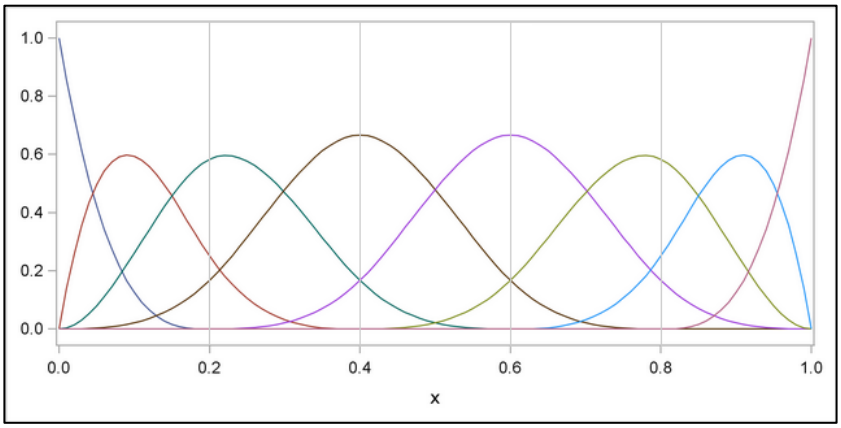
- *Rapidez en el cálculo de los coeficientes* $c_k$*.* Esto son calculados fácilmente resolviendo un sistema de ecuaciones lineales con una ecuación por cada función base.

  Aún si necesitamos de miles de funciones bases no habrá problema alguno al contrario de los polinomios, series de Fourier, ondas las cuales solo funcionan bajo ciertas condiciones como observaciones igualmente espaciadas.

- *Diferenciabilidad:* Las derivadas juegan un papel importante en la modelación de las curvas por lo que deseamos que las funciones bases tengan derivadas que sean suaves. Las series de Fourier trabajan muy bien aquí pero los polinomios disminuyen significativamente su explicación dado que sus derivadas se ven afectadas rápidamente y en la práctica no es así.

- *Restricciones especiales:* Algunas funciones solo pueden ser positivas, algunas otras son monótonas crecientes o decrecientes. Estas restricciones pueden satisfacerse por medio de b-splines de tal manera que las funciones cumplan con las condiciones deseadas.

**Gráfico 2.3:** Funciones bases cúbicas "B-splines" para el intervalo [0,1]. Nótese que la base es conformada por ocho funciones bases definidas por cuatro nudos interiores.




**Fuente:** SAS/STAT(R) 9.2 User's Guide, Second Edition. [12]

## *2.3 Ajuste de los datos usando un sistema base por mínimos cuadrados*

La función $x(t)$ puede ser escrita como una combinación lineal de funciones bases denotado como la expresión

$$x(t) = \sum_{k=1}^{K} c_k \phi_k(t) = \mathbf{c}'\mathbf{\Phi}, \tag{2.8}$$

siendo el vector $\mathbf{c}$ quien contiene los $K$ coeficientes $c_k$ y la matriz $\mathbf{\Phi}$ de dimensiones $n$ por $K$ que contiene los valores de $\phi_k(t_j)$ . Un suavizador lineal simple es obtenido si determinamos los coeficientes de la expansión $c_k$ por el criterio de minimización de mínimos cuadrados

$$\mathrm{SSE}(\mathbf{y}|\mathbf{c}) = \sum_{j=1}^{n}\left[y_j - \sum_{k}^{K} c_k \phi_k(t_j)\right]^2 \qquad (2.9)$$

donde $(2.9)$ escrito de manera matricial se presenta de manera simple como

$$\mathrm{SSE}(\mathbf{y}|\mathbf{c}) = (\mathbf{y} - \mathbf{\Phi c})'(\mathbf{y} - \mathbf{\Phi c}) = \|\mathbf{y} - \mathbf{\Phi c}\|^2. \qquad (2.10)$$

Tomando la derivada parcial de $\mathrm{SSE}(\mathbf{y}|\mathbf{c})$ con respecto al vector $\mathbf{c}$, e igualando a cero podemos calcular el vector estimado $\hat{\mathbf{c}}$ que minimiza el criterio de mínimos cuadrados; esto es:

$$\frac{\partial}{\partial \mathbf{c}} \mathrm{SSE}(\mathbf{y}|\mathbf{c}) = \mathbf{0}$$

$$2\mathbf{\Phi}'\mathbf{\Phi c} - 2\mathbf{\Phi}'\mathbf{y} = \mathbf{0}$$

$$\hat{\mathbf{c}} = \left(\mathbf{\Phi}'\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}'\mathbf{y}.$$

donde el vector $\hat{\mathbf{y}}$ de valores ajustados es

$$\hat{\mathbf{y}} = \mathbf{\Phi}\hat{\hat{\mathbf{c}}} = \mathbf{\Phi}\left(\mathbf{\Phi}'\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}'\mathbf{y}. \qquad (2.11)$$

La aproximación por mínimos cuadrados es apropiada en situaciones donde asumimos que los errores $\epsilon_j$ alrededor de la verdadera curva son i.i.d. con media 0 y varianza constante $\sigma^2$.

## *2.4 Ajuste de los datos usando un sistema base por mínimos cuadrados ponderados*

Para mínimos cuadrados ponderados incluimos a la matriz inversa de varianzas y covarianzas del error $\boldsymbol{\Sigma_e}$ como ponderador esto es

$$\mathbf{W} = \boldsymbol{\Sigma}_{\mathbf{e}}^{-\mathbf{1}}. \tag{2.12}$$

Si $\boldsymbol{\Sigma_e}$ es desconocida puede ser estimada por

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{e}} = (\mathbf{N} - \mathbf{1})^{-\mathbf{1}}\, \mathbf{E}'\mathbf{E}, \tag{2.13}$$

donde $\mathbf{E}$ es la matriz de varianzas y covarianza de los residuos. Por lo tanto la suma cuadrática del error estará dada por

$$\mathrm{SSE}(\mathbf{y}|\mathbf{c}) = (\mathbf{y} - \boldsymbol{\Phi}\mathbf{c})'\mathbf{W}(\mathbf{y} - \boldsymbol{\Phi}\mathbf{c}), \tag{2.14}$$

siendo $\hat{\mathbf{c}}$ el vector de coeficientes que minimiza $(2.14)$ calculado por

$$\hat{\mathbf{c}} = \left(\boldsymbol{\Phi}'\mathbf{W}\boldsymbol{\Phi}\right)^{-\mathbf{1}}\boldsymbol{\Phi}'\mathbf{W}\,\mathbf{y}.$$

El vector $\hat{\mathbf{y}}$ de valores ajustados será

$$\hat{\mathbf{y}} = \boldsymbol{\Phi}\hat{\mathbf{c}} = \boldsymbol{\Phi}\left(\boldsymbol{\Phi}'\mathbf{W}\boldsymbol{\Phi}\right)^{-\mathbf{1}}\boldsymbol{\Phi}'\mathbf{W}\,\mathbf{y} = \mathbf{S}_{\boldsymbol{\Phi}}\mathbf{y}, \tag{2.15}$$

donde $\mathbf{S}_{\boldsymbol{\Phi}}$ es la matriz “hat” u operador de proyección correspondiente al sistema base $\boldsymbol{\Phi}$. La matriz “hat” permite para visualizar los valores estimados de $\mathbf{y}$ como combinaciones lineales de los valores observados [16]. Estos se muestran de la forma

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \hat{\mathbf{y}} = \mathbf{S}_{\boldsymbol{\Phi}}\mathbf{y} = \begin{bmatrix} s_{11}s_{12}s_{13}\cdots s_{1n} \\ s_{21}s_{22}s_{23}\cdots s_{2n} \\ \vdots\ \vdots\ \vdots\ \cdots\ \vdots \\ s_{n1}s_{n2}s_{n3}\cdots s_{nn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \tag{2.16}$$

además se puede probar que la matriz $\mathbf{S}_{\boldsymbol{\Phi}}$ es idempotente es decir que

$$\mathbf{S}_{\boldsymbol{\Phi}}\mathbf{S}_{\boldsymbol{\Phi}} = \mathbf{S}_{\boldsymbol{\Phi}}{}^{\mathbf{2}} = \mathbf{S}_{\boldsymbol{\Phi}}.$$

## *2.5 Derivadas*

Al referirse a una función suave, significa que esta posee una o más derivadas, denotadas por $Dx$, $D^2x$, así hasta $D^m x$, que representa la derivada de orden $m$, donde $D^m x(t)$ es el valor de la derivada de orden m en el instante $t$. Modelar la razón de cambio de un sistema es usualmente llamada análisis de dinámica [11].

La curvatura de una función $x$ en el instante $t$ generalmente es medido por el tamaño de la segunda derivada, calculada como $|D^2x(t)|$ o $[D^2x(t)]^2$. En lugares donde la curvatura es grande, es esencial tener suficientes puntos para estimar la función de manera efectiva.

La cantidad de puntos va a depender de la cantidad de error $\epsilon_j$ definido en la sección 2.1, donde de ser pequeño el error, unas pocas observaciones serían suficientes para hacer inferencia estadística acerca de la aceleración de un sistema. Véase Press et al. [10].

La derivada de primer orden puede ser aproximada por medio de la primera diferencia central denotada como

$$Dx(t_j) \approx \frac{y_{j+1} - y_{j-1}}{t_{j+1} - t_{j-1}}; \qquad (2.17)$$

así mismo, la segunda derivada puede aproximada por la segunda diferencia central denotada por

$$D^2x(t_j) \approx \frac{y_{j+1} + y_{j-1} - 2y_j}{(\Delta t)^2}. \tag{2.18}$$

## *2.6 Cuantificación de aspereza (Roughness)*

Una manera sencilla de cuantificar la aspereza de una función es el cuadrado de la segunda derivada de una función para todo instante $t$, esto es $[D^2x(t)]^2$, lo que usualmente es llamada medida de curvatura en el instante $t$.

Si tenemos una línea recta, dado que no tiene curvatura su segunda derivada siempre será siempre será cero. La aspereza se medirá por

$$\mathrm{PEN}_2(x) = \int [D^2x(t)]^2 dt, \tag{2.19}$$

donde se lo denota $\mathrm{PEN}$ por "*penalization"* que significa penalización en inglés. Si se desea cuantificar la penalización de aspereza para curvas de aceleración, de hecho estas son segundas derivadas, por lo que la penalización para las curvas sería

$$\mathrm{PEN}_4(x) = \int [D^4x(t)]^2 dt\,; \tag{2.20}$$

con lo anteriormente mencionado, podríamos generalizar la penalización de aspereza de orden m como

$$\mathrm{PEN}_m(x) = \int [D^m x(t)]^2 dt. \tag{2.21}$$

Partiendo de la expresión (2.20), se puede expresar la penalización de aspereza $\mathrm{PEN}_m(x)$ como se muestra a continuación

$$\begin{aligned}\mathrm{PEN}_m(x) &= \int [D^m x(t)]^2 dt \\ &= \int \left[D^m \mathbf{c}'\mathbf{\Phi}(t)\right]^2 dt \\ &= \int \mathbf{c}'\, D^m\mathbf{\Phi}(t) D^m\mathbf{\Phi}(t)\, \mathbf{c}\, dt \\ &= \mathbf{c}'\left[\int D^m\mathbf{\Phi}(t) D^m\mathbf{\Phi}(t) dt\right]\mathbf{c} \\ &= \mathbf{c}'\mathbf{R}\,\mathbf{c}\,, \end{aligned} \tag{2.22}$$

donde

$$\mathbf{R} = \int D^m\mathbf{\Phi}(t) D^m\mathbf{\Phi}(t) dt. \tag{2.23}$$

## *2.7 Regularización de Tikhonov*

La regularización de Tikhonov, llamada así por su creador Andrey Tikhonov, es el método de regularización más comúnmente usado para

problemas donde no existe una solución o existen infinitas, lo cual se denota en inglés *ill-posed* [15]. Este método también es llamado método Tikhonov-Miller, método Phillips-Twomey, método de inversión lineal restringido o simplemente método de regularización lineal.

Supongamos que la expresión

$$\mathbf{y} = \mathbf{\Phi c},$$

no está bien condicionada, debido a que no existe solución única o existen infinitas soluciones para $\mathbf{c}$. La aproximación por mínimos cuadrados busca minimizar el residuo

$$\|\mathbf{y} - \mathbf{\Phi c}\|^2,$$

donde ‖·‖ es la norma euclidiana. El no tener solución, puede darse cuando el sistema está sobre-determinado o sub-determinado ($\mathbf{\Phi}$ puede ser una matriz mal condicionada o singular).

Con el fin de dar preferencia a una solución particular con ciertas propiedades deseables, se incluye un término de regularización en esta minimización, esto es

$$\|\mathbf{y} - \mathbf{\phi c}\|^2 + \|\mathbf{\Gamma c}\|^2,$$

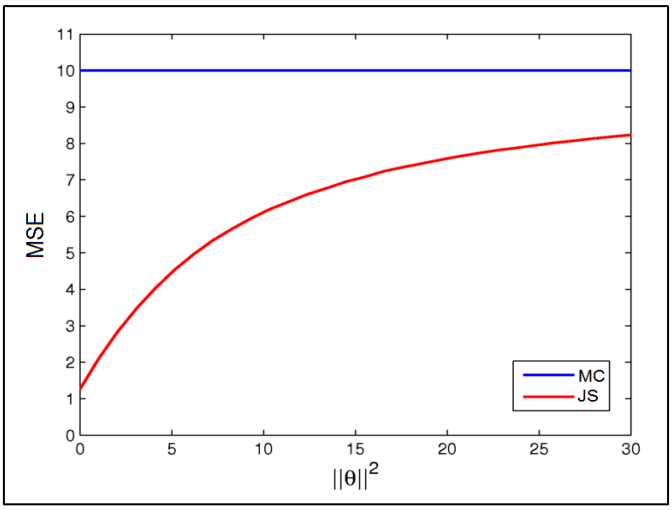
donde $\mathbf{\Gamma}$ es llamada matriz Tikhonov [15]. En algunos casos se toma como la matriz identidad $\mathbf{\Gamma} = \mathbf{I}$ dando lugar a soluciones con normas menores. Esta regularización mejora las condiciones del problema haciendo posible una solución numérica. Una solución explícita

denotada por $\hat{\mathbf{c}}$ es dada por

$$\hat{\mathbf{c}} = (\mathbf{\Phi}'\mathbf{W}\mathbf{\Phi} + \mathbf{\Gamma}'\mathbf{\Gamma})^{-1}\mathbf{\Phi}'\mathbf{W}\,\mathbf{y} \qquad (2.24)$$

donde si $\mathbf{\Gamma}'\mathbf{\Gamma} = \lambda\,\mathbf{\Phi}'\mathbf{\Phi}$, al vector $\hat{\mathbf{c}}$ se lo denomina estimadores de James-Stein. Este estimador cumple la propiedad de tener menor $\mathrm{MSE}$ que el estimador por mínimos cuadrados siempre que $\dim(\hat{\mathbf{c}}) \geq 3$, lo que hace que este estimador sea admisible bajo esta condición [14]. Véase gráfico 2.4.

**Gráfico 2.4:** MSE del estimador de Mínimos Cuadrados (MC) vs. El estimador de James–Stein(JS). El estimador de James–Stein tiene una mejor estimación cuando la norma del parámetro es cercana a cero.




**Fuente:** Wikipedia. James–Stein estimator.

Uno de los casos particulares es la regresión Ridge, donde $\mathbf{\Gamma}'\mathbf{\Gamma} = \lambda\mathbf{I}$ y el vector de estimadores $\hat{\mathbf{c}}$ es

$$\hat{\mathbf{c}} = (\mathbf{\Phi}'\mathbf{W}\mathbf{\Phi} + \lambda\mathbf{I})^{-1}\mathbf{\Phi}'\mathbf{W}\,\mathbf{y}. \qquad (2.25)$$

Es fácil demostrar que al añadir el término $\lambda\mathbf{I}$ a la minimización se puede lograr que la matriz $\boldsymbol{\Phi}'\mathbf{W}\boldsymbol{\Phi}$ sea inversible, llegando a tener una solución particular para $\hat{\mathbf{c}}$, donde además se logra reducir el error cuadrático medio del vector de estimaciones. Para mayor detalle léase Hoerl [4].

En el caso particular del Análisis de Datos Funcionales, $\boldsymbol{\Gamma}'\boldsymbol{\Gamma} = \lambda\mathbf{R}$ donde la matriz $\mathbf{R}$ se encuentra definida en (2.23) y $\lambda$ es el parámetro de suavización. El vector de estimadores es calculado como

$$\hat{\mathbf{c}} = (\boldsymbol{\Phi}'\mathbf{W}\boldsymbol{\Phi} + \lambda\mathbf{R})^{-1}\boldsymbol{\Phi}'\mathbf{W}\,\mathbf{y}\,, \tag{2.26}$$

por lo que $\boldsymbol{\Gamma} = \sqrt{\lambda}\,\mathbf{R}^{1/2}$, donde $\mathbf{R}^{1/2}$ es la matriz raíz cuadrada de la matriz $\mathbf{R}$ anteriormente definida, la cual puede ser calculada por descomposición de Cholesky.

Para este caso particular, la matriz “hat” correspondiente al sistema base $\boldsymbol{\Phi}$, con parámetro de penalización $\lambda$ estará definida por

$$\mathbf{S}_{\boldsymbol{\Phi},\lambda} = \boldsymbol{\Phi}\left(\boldsymbol{\Phi}'\mathbf{W}\boldsymbol{\Phi} + \lambda\mathbf{R}\right)^{-1}\boldsymbol{\Phi}'\mathbf{W}, \tag{2.27}$$

donde a la matriz $\mathbf{S}_{\boldsymbol{\Phi},\lambda}$ se la llama de operador de sub-proyección, dado que a diferencia de los operadores de proyección, no cumple la propiedad de idempotencia, es decir:

$$\mathbf{S}_{\boldsymbol{\Phi},\lambda}\mathbf{S}_{\boldsymbol{\Phi},\lambda} \neq \mathbf{S}^{2}_{\boldsymbol{\Phi},\lambda}.$$

## *2.8 Método de Validación Cruzada Generalizada*

El método de validación cruzada generalizada o método GCV, esto por su acrónimo en lengua inglesa *generalized cross-validation*, fue desarrollado por Craven y Wahba en 1979, siendo una de las medidas más frecuentemente mencionadas en la literatura de suavizamiento, el cual tiene como principal objetivo determinar el valor óptimo del estimador del parámetro de suavización $\lambda$.

Según [11], el mejor estimador será el valor de lambda que minimice la expresión

$$\mathrm{GCV}(\lambda) = \frac{n^{-1}\,\mathrm{SSE}}{\left[n^{-1}\,\mathrm{traza}(\mathbf{I} - \mathbf{S}_{\mathbf{\Phi},\lambda})\right]^2} \tag{2.28}$$

o la expresión equivalente

$$\mathrm{GCV}(\lambda) = \left(\frac{n}{n - \mathrm{df}(\lambda)}\right)\left(\frac{\mathrm{SSE}}{n - \mathrm{df}(\lambda)}\right) \tag{2.29}$$

donde el segundo multiplicando es el estimador insesgado de la varianza del error $\sigma^2$ usualmente conocido en regresión y los grados de libertad son equivalentes a la traza de la matriz $\mathbf{S}_{\mathbf{\Phi},\lambda}$ anteriormente definida en (2.27).

Este método se basa en método de validación cruzada simple y tiene las ventajas de existe una menor tendencia a sobre-suavizar y evita la

necesidad de volver a suavizar para las $n$ ocasiones. Siempre se deberá probar con grandes cantidades de valores de $\lambda$ para minimizar el GCV por lo que es necesario el uso de métodos numéricos por medio de algoritmos de optimización numérica [11].

## *2.9 Suma Cuadrática del Error penalizada de orden $m$*

La Suma Cuadrática del Error penalizada de orden m denotada por $\mathrm{PENSSE}_m$, se obtiene sumando la Suma Cuadrática del Error expresada en (2.14) y la penalización de orden $m$ (2.22); esta última se debe multiplicar por el parámetro de suavización $\lambda$ obtenido previamente por el Método de Validación Cruzada Generalizada. De esta manera tenemos

$$\mathrm{PENSSE}_m(\mathbf{y}|\mathbf{c}) = \mathrm{SSE}(\mathbf{y}|\mathbf{c}) + \lambda\, \mathrm{PEN}_m(\mathbf{x})$$

$$\mathrm{PENSSE}_m(\mathbf{y}|\mathbf{c}) = (\mathbf{y} - \boldsymbol{\Phi}\mathbf{c})'\mathbf{W}(\mathbf{y} - \boldsymbol{\Phi}\mathbf{c}) + \lambda \mathbf{c}'\mathbf{R}\mathbf{c}. \qquad (2.30)$$

Tomando la derivada parcial con respecto al vector de parámetros $\mathbf{c}$, obtenemos

$$\frac{\partial}{\partial \mathbf{c}}(\mathbf{y} - \boldsymbol{\Phi}\mathbf{c})'\mathbf{W}(\mathbf{y} - \boldsymbol{\Phi}\mathbf{c}) + \lambda \mathbf{c}'\mathbf{R}\mathbf{c} = 0$$

$$-2\boldsymbol{\Phi}'\mathbf{W}\mathbf{y} + \boldsymbol{\Phi}'\mathbf{W}\boldsymbol{\Phi}\mathbf{c} + \lambda \mathbf{R}\mathbf{c} = 0,$$

de la cual obtenemos la expresión para el vector de coeficientes estimados

$$\hat{\boldsymbol{c}} = \left(\boldsymbol{\Phi}'\mathbf{W}\boldsymbol{\Phi} + \lambda\mathbf{R}\right)^{-1}\boldsymbol{\Phi}'\mathbf{W}\,\mathbf{y}, \qquad (2.31)$$

expresión que coincide con (2.26) resultante de la Regularización de Tikhonov cuando $\boldsymbol{\Gamma} = \sqrt{\lambda}\,\mathbf{R}^{1/2}$.

## *2.10 Funciones Monótonas*

Usualmente sabemos que en ciertos procesos las mediciones son valores que siempre se están incrementando como por ejemplo la estatura de una persona, es por esto que algunas veces hablamos de datos o funciones monótonas crecientes o decrecientes.

La monotonicidad es un principio global en la suavización, de modo que eliminando los valores negativos para la primera derivada puede estabilizar las curvas, especialmente al inicio y al final de ellas.

En vez de usar una función para directamente ajustar los datos, la utilizaremos para definir otra función que cumpla con las características requeridas, en este caso la monotonicidad [6].

Supongamos que $W(t)$ es una función cualquiera la cual no tiene restricciones excepto de que $W(t_0) = 0$. Entonces $\exp[W(t)]$ es una función siempre positiva y sabemos que la integral definida de una función positiva siempre se incrementa. En consecuencia, nosotros definimos una función positiva $m(t)$ de la forma

$$m(t) = e^{W(t)},$$

donde $W(t)$ es

$$W(t) = \sum_k c_k \phi_k(t),$$

por lo que la expresión a minimizar por mínimos cuadrados ponderados penalizados es

$$\text{PENSSE}_\lambda(W|\mathbf{y}) = \left[\mathbf{y} - e^{W(\mathbf{t})}\right]' \mathbf{W}\left[\mathbf{y} - e^{W(\mathbf{t})}\right] + \lambda \int [D^2 W(t)]^2 dt\,,$$

donde dada la naturaleza de la ecuación, el cálculo computacional se verá complicado en comparación que al proceso de suavización usual [11].

El principio de monotonicidad implica que la primera derivada o velocidad de $x$ sea siempre positiva, lo cual puede ser denotado como una ecuación diferencial, esto es $Dx(t) = e^{W(t)}$, por lo que integrando a ambos lados podemos expresar a la función $x(t)$ como

$$x(t) = C + \int_{t_0}^{t} exp[W(u)]\, du. \qquad (2.28)$$

## *2.11 Factor de Inflación de la Varianza*

El Factor de Inflación de la Varianza o VIF (*Variance Inflation Factor*) mide la cantidad de multicolinealidad que existe en un modelo de regresión por Mínimos Cuadrados Ordinarios.

El *VIF* es un índice del aumento del error estándar dada la multicolinealidad que existen entre las variables independientes del modelo.

Suponiendo el siguiente modelo lineal que tiene $p$ variables

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p X_p + \varepsilon, \tag{2.29}$$

la matriz de varianzas y covarianzas de los estimadores de regresión $\mathbf{\Sigma_b}$ es estimada de manera matricial por $\mathrm{MCE}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$ [16], mas el estimador de la varianza de cada uno de los estimadores $\beta_j$, según Kutner [9] puede ser calculado como

$$\widehat{var}(\widehat{\beta}_j) = \frac{\mathrm{MCE}}{(n-1)\widehat{var}(x_j)} \left(\frac{1}{1-R_j^2}\right), \tag{2.30}$$

donde $R_j^2$ es el coeficiente de determinación $R^2$ para el modelo de regresión donde $x_j$ es explicado por medio de las otras covariables, esto es

$$x_j = \gamma_0 + \cdots + \gamma_{j-1} x_{j-1} + \gamma_{j+1} x_{j+1} + \cdots + \xi. \tag{2.31}$$

El segundo factor, $1 / 1 - R_j^2$ es el VIF, el cual indica en cuantas veces se verá incrementado el error estándar para el estimador $\widehat{\beta_j}$ a causa de la multicolinealidad. Si el VIF es igual a uno, esto implica que $x_j$ es linealmente independiente con las restantes covariables, lo que matricialmente representaría que el vector $x_j$ es ortogonal con respecto a las columnas correspondientes a las otras covariables en la matriz de diseño.

Kutner [9] sugiere que si el VIF es mayor a diez, se podría considerar que existe una alta multicolinealidad, lo que sugeriría que $x_j$ puede ser excluida del modelo. El método VIF se encuentra debidamente implementado en el software estadístico R como `VIF()` y se encuentra en la librería `car`.

# CAPÍTULO III

## 3.1. ANÁLISIS DE VARIANZA FUNCIONAL

### *3.1.1. Introducción*

El Análisis de Varianza (ANOVA) es una de las herramientas más usadas en la estadística aplicada, esto es para el Diseño de Experimentos. Mientras esta es muy útil cuando se trabaja con datos de pequeñas dimensiones, tiene sus limitaciones analizando variables de interés funcionales.

Tales variables funcionales son encontradas, por ejemplo, cuando las unidades son tomadas alrededor del tiempo, incluso cuando no se tiene en sí una función pero si un conjunto de evaluaciones individuales suficientes para suponerla [1].

En tales casos, el método de Análisis de Varianza Funcional (FANOVA) provee igual solución que el tradicional ANOVA cuando la variable dependiente es funcional, sin dejar de lado su poder de explicación y facilidad de interpretación.

### *3.1.2. Aplicación del Análisis Funcional de Varianza*

Se desea determinar si existen diferencias significativas en ciertas características físico químicas de las papayas, al usar dos tipos diferentes de películas comestibles, por lo que la metodología a usarse es el análisis de varianza funcional dado que la variable dependiente constituye un dato funcional.

En términos formales, si $Y$ es la característica a medir en la papaya y tenemos $P$ tipos de películas comestibles, el modelo que explicaría la $r-ésima$ observación funcional con el tipo de película $p$ denotada por $Y_{pr}$ sería

$$Y_{pr}(t) = \beta_0(t) + \alpha_p(t) + \epsilon_{pr}(t) \qquad (3.1)$$

donde $\beta_0(t)$ es la función promedio $\mu(t)$ de la característica de interés $Y$ para papayas sin película, $\alpha_p(t)$ es el efecto sobre la característica $Y$ al usar el tipo de película $p$. El término $\epsilon_{pr}(t)$ representa la función residual, la cual es la variación no explicada respecto a la $r-ésima$ papaya con película comestible $p$. Nótese que es de interés analizar $\alpha_1(t)$ y $\alpha_2(t)$ dada la condición que

$$\alpha_0(t) = 0, \qquad \forall t. \qquad (3.2)$$

Podemos definir las tres funciones de regresión $\beta_j$ correspondientes a

las constantes del modelo anterior, esto sería $\beta_1 = \alpha_1$ y $\beta_2 = \alpha_2$. De esta manera el modelo puede ser expresado por

$$Y(t) = \beta_0(t) + \beta_1(t)Z_1 + \beta_2(t)Z_2 + \epsilon(t); \tag{3.3}$$

también se puede escribir (3.3) de manera general [11] como

$$Y_{pr}(t) = \sum_{j}^{p} Z_{(pr)j}\,\beta_j(t) + \epsilon_{pr}(t), \tag{3.4}$$

o bien podríamos expresarlo de manera compacta en forma matricial por medio de la expresión

$$\mathbf{Y} = \mathbf{Z\beta} + \boldsymbol{\epsilon}, \tag{3.5}$$

donde $\mathbf{Z}$ es la matriz de diseño y $\boldsymbol{\beta}' = [\beta_0, \beta_1, \beta_2]$ ambos bajo la condición en (3.2).

### *3.1.3. Ajuste del Modelo*

Si tuviéremos un modelo lineal estándar se debería usar el método de Mínimos Cuadrados para hallar el vector $\boldsymbol{\beta}$ que minimice la Suma Cuadrática Residual. Extendiendo el criterio de Mínimos Cuadrados de manera adecuada para el caso funcional, tenemos que cada uno de los residuos $Y_{pr}(t) - \mathbf{Z}_{(pr)}\boldsymbol{\beta}(t)$ constituyen una función, por lo que el criterio de ajuste por Mínimos Cuadrados no ponderados sería

$$\mathrm{LMSSE}(\boldsymbol{\beta}) = \sum_{p}\sum_{r}\int_{t_0}^{t_f}\left[Y_{pr}(t) - \sum_{j=0}^{P} Z_{(pr),j}\beta_j(t)\right]^2 dt, \qquad (3.6)$$

donde $t_0$ y $t_f$ representan el tiempo inicial y final de la realización del experimento respectivamente. Téngase en cuenta, que la minimización en realidad es respecto a el vector de coeficientes $\mathbf{c}$, dado que los $\beta_j(t)$ son combinaciones lineales de los elementos del sistema base $\boldsymbol{\Phi}$.

Bajo la restricción que $Z_{(pr),0} = 1$ se obtiene la estimación del vector de parámetros $\boldsymbol{\beta}$ por Mínimos Cuadrados, siendo la medida cuantitativa del error el área bajo la curva de las funciones residuales al cuadrado.

### *3.1.4. Validación del Modelo*

Para la validación del modelo al igual que en el análisis de varianza común utilizamos el Coeficiente de Determinación $R^2$ y el Estadístico de prueba $F$ de Fisher. Para el caso de Análisis de Varianza Funcional, $R^2$ y $F$ son funciones, al igual que las sumas y medias cuadráticas.

Al igual que en los modelos lineales multivariados, la principal fuente de información en el análisis de la importancia de los efectos es la Suma Cuadrática del Error que sería

$$\mathrm{SSE}(t) = \sum_{rp} \left[ Y_{pr}(t) - \mathbf{Z}_{\mathrm{pr}} \widehat{\boldsymbol{\beta}}(\mathrm{t}) \right]^2 \tag{3.7}$$

La Suma Cuadrática Total, puede ser comparada como la Suma Cuadrática del Error cuando en el modelo solo se considera la media global $\hat{\mu}$ como predictor. Esto es,

$$\mathrm{SST}(t) = \sum_{rp} \left[ Y_{pr}(t) - \hat{\mu}(t) \right]^2 \tag{3.8}$$

donde la media global $\hat{\mu}(t)$ puede ser hallada de manera descriptiva como en (1.1), o dado que el diseño es balanceado puede ser calculada en función de los efectos como

$$\hat{\mu}(t) = \frac{3\hat{\beta}_0(t) + \hat{\beta}_1(t) + \hat{\beta}_2(t)}{3}, \tag{3.9}$$

La Potencia de Explicación del Modelo puede ser calculada por medio de la función de Correlación Múltiple Cuadrada $\mathrm{RSQ}$ análoga del coeficiente $R^2$ usado en Análisis de Varianza Simple, esta es calculada como

$$\mathrm{RSQ}(t) = \frac{[\mathrm{SST}(t) - \mathrm{SSE}(t)]}{\mathrm{SST}(t)}, \tag{3.10}$$

donde el numerador corresponde a la función Suma Cuadrática de

Regresión. De igual manera podemos calcular las funciones análogas correspondientes a los estadísticos de la tabla ANOVA para el análisis univariado. Por ejemplo, la función Media Cuadrática del Error tiene la forma funcional,

$$\mathrm{MSE}(t) = \frac{\mathrm{SSE(t)}}{\mathrm{df}_{(error)}}, \tag{3.11}$$

donde $\mathrm{df}_{(error)}$ son los grados del libertad del error correspondientes al tamaño de la muestra menos el número de funciones independientes $\beta_i$ en el modelo. La función Media Cuadrática de Regresión sería

$$\mathrm{MSR}(t) = \frac{\mathrm{SSY}(t) - \mathrm{SSE}(t)}{\mathrm{df}_{(reg)}}. \tag{3.12}$$

Finalmente podemos calcular la función análoga al estadístico $F$ de Fisher llamado en literatura funcional como $\mathrm{F}_{RATIO}$ la cual se calcula como

$$\mathrm{F}_{RATIO}(t) = \frac{\mathrm{MSR}(t)}{\mathrm{MSE}(t)}, \tag{3.13}$$

y con esto hallar la función $\mathrm{P}_{VALUE}$ equivalente al valor p correspondiente a la Prueba $F$, la cual nos permitirá concluir si en algún instante de tiempo $t$ durante el experimento, existe algún efecto significativo sobre la variable de interés $Y$ al aplicar un cierto tipo de película. Obsérvese en la tabla 1, el compendio de funciones anteriormente definidas resumidas como la tabla FANOVA.

**Tabla 1:** Tabla de Análisis de Varianza Funcional (FANOVA)

| Fuentes de Variación | Grados de Libertad | Sumas Cuadráticas | Medias Cuadráticas | Estadístico de Prueba F |
|---|---|---|---|---|
| REGRESION | $p-1$ | $\sum_{rp}\left[\mathbf{Z}_{\mathrm{pr}}\widehat{\boldsymbol{\beta}}(\mathrm{t})-\hat{\mu}(t)\right]^2$ | $\frac{\mathrm{SSR}(t)}{\mathrm{df}_{(regresión)}}$ | $\frac{\mathrm{MSR}(t)}{\mathrm{MSE}(t)}$ |
| ERROR (Residuales) | $n-p$ | $\sum_{rp}\left[Y_{pr}(t)-\mathbf{Z}_{\mathrm{pr}}\widehat{\boldsymbol{\beta}}(\mathrm{t})\right]^2$ | $\frac{\mathrm{SSE(t)}}{\mathrm{df}_{(error)}}$ | |
| TOTAL | $n-1$ | $\sum_{rp}\left[Y_{pr}(t)-\hat{\mu}(t)\right]^2$ | | |

Bajo las condiciones dadas, el contraste de hipótesis respectivo para la prueba F de FANOVA sería:

$$H_0\colon \alpha_1(t)=\alpha_2(t)=0$$

$$vs$$

$$H_1\colon Al\ menos\ una\ función\ \alpha(t)\neq 0$$

Es decir la hipótesis nula postula que los efectos de los tratamientos son nulos, que es lo mismo decir, sin importar el tipo de película que se aplique, en promedio siempre se tendrá iguales resultados. Con un $(1-\alpha)100\%$ de confianza la hipótesis nula $H_0$ será rechazada en favor de $H_1$ en el instante de tiempo $t$ si

$$\mathrm{F}_{RATIO}(t) > F_{(\alpha;\,\mathrm{df}_{(reg)},\mathrm{df}_{(error)})}, \qquad (3.14)$$

donde $F_{(\alpha;\,\mathrm{df}_{(reg)},\mathrm{df}_{(error)})}$ es el percentil $(1-\alpha)100$ de la distribución F con $\mathrm{df}_{(reg)}$ grados de libertad en el numerador y $\mathrm{df}_{(error)}$ grados de libertad en el denominador. Si deseamos rechazar en base al valor p, llamaremos $\mathrm{P}_{VALUE}$ a su función análoga la cual puede ser calculada como

$$\mathrm{P}_{VALUE}(t) = P\left(F_{(\mathrm{df}_{(reg)},\mathrm{df}_{(error)})} \geq \mathrm{F}_{RATIO}(t)\right). \qquad (3.15)$$

## 3.2. PRUEBA T FUNCIONAL PARA DIFERENCIA DE MEDIAS PARA POBLACIONES NORMALES CON VARIANZA DESCONOCIDA

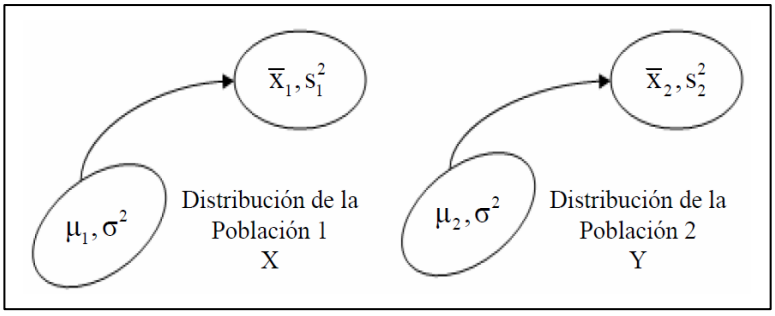
Será de interés determinar si el uso de aceites esenciales en las películas comestibles tiene un efecto deseable sobre alguna de las características de interés $Y$. Con este fin, se realizarán pruebas de hipótesis funcionales para diferencias de medias, comparando los dos tipos de películas, cuando sea necesario.

Para esto, se debe tener dos muestras independientes bajo la condición, que ambas poblaciones de las que se toman las muestras son Normales e independientes. Nótese que no se exigen las condiciones del Teorema del Límite Central [16].

### 3.2.1 *Prueba T funcional para diferencia de Medias para Poblaciones Normales con Varianzas desconocidas e iguales*

Bajo el supuesto que en un instante $t$ del experimento, $X_1 \sim N(\mu_1(t), \sigma^2(t))$ mientras que $X_2 \sim N(\mu_2(t), \sigma^2(t))$ siendo $\sigma_1^2(t) = \sigma_2^2(t) = \sigma^2(t)$, se desconoce $\sigma^2(t)$; $n_1$ y $n_2$ serán los tamaños de muestra correspondientes a las muestras independientes $\mathbf{X_1}$ y $\mathbf{X_2}$ los cuales permanecen constantes a lo largo del experimento [16].

**Gráfico 3.1:** Poblaciones Normales con igual varianza. No se exigen las condiciones del Teorema del Límite Central.




**Fuente:** Probabilidad y Estadística, Fundamentos y Aplicaciones, Zurita [16]

Dado el supuesto que las varianzas iguales para todo $t$, se calcula un Estimador de la función varianza común $\sigma^2(t)$ por medio de las funciones de varianzas muestrales $s_1^2(t)$ y $s_2^2(t)$. Este estimador es análogo al estimador S *polled* para el caso no funcional, y es

calculado como

$$s_p^2(t) = \frac{(n_1 - 1)s_1^2(t) + (n_2 - 1)s_2^2(t)}{n_1 + n_2 - 2}. \quad (3.16)$$

El estadístico de prueba funcional es

$$T(t) = \frac{\bar{x}_1(t) - \bar{x}_2(t) - \delta(t)}{s_p(t)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad (3.17)$$

donde $\delta(t)$ es la función correspondiente al valor de la diferencia de medias postulada en la hipótesis nula. El estadístico $T(t)$ tiene distribución T de Student con $n_1 + n_2 - 2$ grados de libertad para todo $t$.

Bajo las condiciones dadas, el contraste de hipótesis unilateral respectivo para la prueba funcional T sería

$$H_0: \mu_1(t) - \mu_2(t) \geq 0$$

$$vs$$

$$H_1: \mu_1(t) - \mu_2(t) < 0.$$

La hipótesis nula postula que ambas funciones promedio son iguales durante todo el experimento. Con un $(1 - \alpha)100\%$ de confianza la hipótesis nula $H_0$ será rechazada en favor de $H_1$ en el instante de tiempo $t$ si

$$T(t) < -T_{(\alpha\,; n_1 + n_2 - 2)}, \quad (3.18)$$

donde $T_{(\alpha\,;n_1+n_2-2)}$ es el percentil $(1-\alpha)100$ de la distribución T con $n_1+n_2-2$ grados de libertad. Para este caso función $\mathrm{P}_{VALUE}(t)$ es calculada como

$$\mathrm{P}_{VALUE}(t) = P\left(T_{(n_1+n_2-2)} \leq T(t)\right). \qquad (3.19)$$

### *3.2.2 Prueba T funcional para diferencia de Medias para Poblaciones Normales con Varianzas desconocidas y desiguales*

Bajo el mismo supuesto de normalidad para las poblaciones $\mathrm{X}_1$ y $\mathrm{X}_2$, en todo instante $t$ del experimento, además $\sigma_1^2(t)$ y $\sigma_2^2(t)$ son desconocidas y se suponen desiguales; $n_1$ y $n_2$ serán los tamaños de muestra correspondientes a las muestras independientes $\mathbf{X_1}$ y $\mathbf{X_2}$ los cuales para este experimento permanecen constantes durante su realización.

Bajo los supuestos anteriores, y siendo $\delta(t)$ la función correspondiente al valor de la diferencia de medias postulada en la hipótesis nula, el Estadístico de Prueba funcional $T(t)$ puede ser calculado como

$$T(t) = \frac{\bar{x}_1(t) - \bar{x}_2(t) - \delta}{\sqrt{\frac{s_1^2(t)}{n_1} + \frac{s_2^2(t)}{n_2}}}, \tag{3.20}$$

tiene distribución T de student con $\nu(t)$ grados de libertad, los cuales son calculados como

$$\nu(t) = \frac{\left(\frac{s_1^2(t)}{n_1} + \frac{s_2^2(t)}{n_2}\right)^2}{\frac{\left(\frac{s_1^2(t)}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2(t)}{n_2}\right)^2}{n_2 - 1}}; \tag{3.21}$$

donde es el estadístico de prueba anterior y su aproximación para los grados de libertad, es conocida en literatura estadística como "*Satterwaite's Test*" [13]. Como los valores de $\nu(t)$ pueden ser decimales, estos deben ser redondeados. Nótese que los grados de libertad $\nu(t)$, y del Estadístico de Prueba $T(t)$ constituyen también una función.

Bajo las condiciones dadas, el contraste de hipótesis unilateral respectivo para la prueba funcional T sería

$$H_0: \mu_1(t) - \mu_2(t) \geq 0$$

$$vs$$

$$H_1: \mu_1(t) - \mu_2(t) < 0,$$

donde la hipótesis nula postula que ambas funciones promedio son iguales durante todo el experimento. Con un $(1-\alpha)100\%$ de

confianza la hipótesis nula $H_0$ será rechazada en favor de $H_1$ en el instante de tiempo $t$ si

$$T(t) < -T_{(\alpha\,;\nu(t))}, \tag{3.22}$$

donde $T_{(\alpha\,;\nu(t))}$ es el percentil $(1-\alpha)100$ de la distribución T con $\nu(t)$ grados de libertad. Para este caso función $\mathrm{P}_{VALUE}(t)$ es calculada como

$$\mathrm{P}_{VALUE}(t) = P\left(T_{(\nu(t))} \leq T(t)\right). \tag{3.23}$$

## 3.3. IMPUTACIÓN DE DATOS

### *3.3.1. Justificación*

Existen características de interés que requieren que el fruto sea cortado para poder ser medidas, es por esto que $r$ papayas seleccionadas aleatoriamente deberán ser cortadas cada día con este objetivo, por lo que se tendrán mediciones faltantes para características que son simplemente observables, por lo que será necesario utilizar un método de imputación de datos para poder estimar estos valores.

### *3.3.2. Modelo de Imputación*

Básicamente, las observaciones faltantes $O_i$ serán calculadas utilizando la función media $\bar{y}_i$, de tal manera que restándole una función lineal $f_i$ cuyos coeficientes $\beta_0$ y $\beta_1$ son de desconocidos y de interés, se ajuste a los datos observados de mejor manera por medio del criterio de Mínimos Cuadrados, esto es:

$$O_i = \bar{y}_i - f_i + \epsilon_i, \tag{3.24}$$

donde $f_i$ es una función lineal de la forma

$$f_i = \beta_0 + \beta_1 t_i, \tag{3.25}$$

donde $\epsilon_i$ es el error, el cual se supone que se distribuye normalmente con media $0$ y varianza constante $\sigma^2$, esto es $\epsilon_i \sim \mathrm{N}(0, \sigma^2)$, además que que $\mathrm{cov}(\epsilon_i, \epsilon_j) = 0$ para $\mathrm{i} \neq \mathrm{j}$. Dada la expresión (3.24), la Suma Cuadrática del Error estará dada por la expresión

$$Q = \sum_{i=1}^{t} [\epsilon_i^2] = \sum_{i=1}^{t} [O_i - \bar{y} + f_i]^2, \tag{3.26}$$

donde reemplazando $f_i$ como en la expresión (3.25) tenemos

$$\sum_{i=1}^{t} [O_i - \bar{y} + \beta_0 + \beta_1 t_i]^2. \tag{3.27}$$

Multiplicando por $(-1)$ dentro del paréntesis y agrupando de manera adecuada, puede expresar a (3.27) como

$$\sum_{i=1}^{t}[(\bar{y} - O_i) - (\beta_0 + \beta_1 t_i)]^2, \tag{3.28}$$

donde, si llamamos

$$Y_i = (\bar{y} - O_i), \tag{3.29}$$

la expresión resultante será

$$\sum_{i=1}^{t}[Y_i - (\beta_0 + \beta_1 t_i)]^2, \tag{3.30}$$

siendo la expresión de la Suma Cuadrática del Error a minimizar por el Método de Regresión Lineal Simple [16].

Los estimadores $b_0$ y $b_1$ que minimizan (3.26) pueden ser hallados por las Ecuaciones Normales, que forman el sistema de ecuaciones

$$\begin{cases} \sum_{i=1}^{t} y_i = tb_0 + \sum_{i=1}^{t} t_i \\ \sum_{i=1}^{t} t_i\, y_i = b_0 \sum_{i=1}^{t} t_i + b_1 \sum_{i=1}^{t} t_i \end{cases}, \tag{3.26}$$

bajo la condición $Y_i = (\bar{y} - O_i)$.

### *3.3.3. Algoritmo de Imputación*

El *algoritmo de imputación* fue programado en el software estadístico R haciendo uso de la librería “fda”. Obsérvese el código en la Tabla 2.

**Tabla 2:** Algoritmo de Imputación en R para observaciones faltantes.

```
i=complete.cases(t(O))
inot=!i
O.comp=O[,i]
O.notcomp=O[,inot]
base=create.basis(range(tiempo),nbasis)
Ofd.comp=apply(O.comp,2,function(y)
      Data2fd(tiempo[!is.na(y)],y[!is.na(y)],base))
O.comp.media[j]=eval.fd(Ofd.comp,tiempo)
Y=matrix(nrow=length(tiempo),ncol=ncol(O.notcomp))
B=matrix(nrow=2,ncol=ncol(O.notcomp))
lengthy=vector(length=ncol(O.notcomp))
F=matrix(nrow=nrow(O.notcomp),ncol=ncol(O.notcomp))
Ob=matrix(nrow=nrow(F),ncol=ncol(F))
for(i in 1:ncol(O.notcomp))
{
  for(j in 1:length(tiempo))
  {
  Y[j,i]= O.comp.media[j]-O.notcomp[j,i]
  }
  reg=lm(Y[,i]~tiempo)
  B[1,i]=reg$coefficients[1]
  B[2,i]=reg$coefficients[2]
  lengthy[i]=length(na.omit(Y[,i]))
}
B=replace(B,is.na(B),0)
for(j in 1:ncol(Z))
{
  for(i in 1:lengthy[j])
{
    F[i,j]=B[1,j]+B[2,j]*tiempo[i]
  }
  F[,j]=na.locf(F[,j])
  Ob[,j]= O.comp.media[j]-F[,j]
}
Obdata=matrix(c(Ob,O.comp),nrow=n)
colnames(Obdata)=c(colnames(O.notcomp),colnames(O.comp)
```

# CAPÍTULO IV

## 4. EXPERIMENTACIÓN

El principal objetivo del experimento es medir si el uso de películas comestibles en papayas puede retrasar significativamente el proceso de madurez. El estado de madurez de la fruta al ser cosechados, es especialmente importante para su manejo, transportación y comercialización ya que repercute directamente en su calidad y potencial de almacenamiento [7].

Por esto, el análisis se centra en ciertas características físico químicas de interés para el consumidor, y que pueden ser indicadores de la maduración del producto.

### *4.1. Variables*

#### *Variable independiente*

La variable independiente o factor a considerarse fue el tipo de película comestible a aplicarse. Esta variable categórica toma tres posibles

valores que son:

- ***Placebo:*** Es la no aplicación de película. También llamado control, este tiene como función medir los efectos de la aplicación de las películas a experimentarse.

- ***Película comestible estándar:*** Película comestible elaborada de almidón de maíz la cual actúa como barrera externa a agentes externos como humedad, aceites, vapor, etc.

- ***Película comestible con aceites esenciales:*** Película comestible estándar con aceites esenciales de canela y clavo de olor, con el fin de controlar de las pudriciones fungosas.

### *Variables de Interés*

Fueron de interés las características que son percibidas por el consumidor como “sinónimo” de calidad y que al mismo tiempo, reflejan la etapa de maduración y próximo perecimiento de una fruta. Las variables de interés son: peso, nivel de acidez pH, sólidos solubles, dureza y color. La descripción de cada una de las

características y la metodología de medición se detalla a continuación:

- ***Peso***

  Todas las papayas existentes al día *t* fueron pesadas (gramos) en una báscula de precisión con capacidad de 2.5 Kg. Esta fue escogida como variable de diseño dada su influencia al momento de adquirir el fruto, facilidad de medición y por ser teóricamente la de mayor expectativa a ser afectada.

- ***Nivel de Acidez***

  Influyente en el sabor que se percibe de la fruta, el nivel de acidez o basicidad en la fruta fue medida en escala pH. Se homogenizaron 10 gramos de pulpa de papaya en 100 ml de agua destilada, y se procedió a medir el nivel pH a temperatura ambiente con un potenciómetro.

- ***Sólidos solubles***

  Los sólidos solubles totales en una fruta están formados en más del 90% por azúcar y el resto por ácidos orgánicos y minerales por lo que esta característica es de vital importancia sobre la dulzura de la papaya. Se mide en grados brix (°brix), analizando

pequeñas muestras de jugo extraído sobre un prisma de medición mediante un refractómetro de mano.

- ***Dureza***

  Se determinó utilizando un penetrómetro manual, introduciendo su cilindro metálico sobre una de las caras en la zona media de cada papaya; la dureza se mide en términos de la presión ejercida sobre la fruta, expresada en $g/cm^2$.

- ***Color***

  Característica física de gran interés la cual fue medida por medio de un colorímetro el cual descomponía el color en coordenadas RGB (Red, Green, Blue).

**Tabla 3:** Colores básicos en sistema de color RGB.

| Color | [ R , G , B ] | Muestra de Color |
|---|---|---|
| Negro | [ 0 , 0 , 0 ] | |
| Rojo | [ 255 , 0 , 0 ] | |
| Verde | [ 0 , 255 , 0 ] | |
| Azul | [ 0 , 0 , 255 ] | |
| Blanco | [ 255 , 255 , 255 ] | |

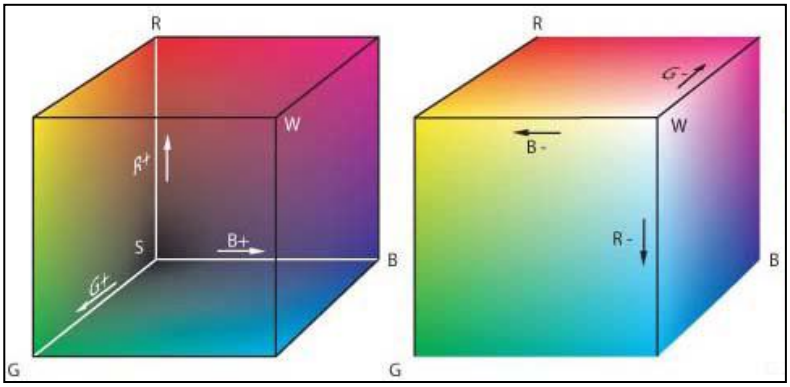
En las coordenadas RGB, cada una de sus componentes puede tomar valores entre 0 y 255, lo que da lugar a $256^3$ posibles colores. Los colores base se componen como se muestra en la tabla 3.

Al tener un vector trivariado, se utilizó la Norma Euclidiana o módulo del vector como cuantía de la cantidad de color, esto es

$$\|V\| = \sqrt{\mathrm{R}^2 + \mathrm{G}^2 + \mathrm{B}^2}, \qquad (4.1)$$

que gráficamente representa la distancia entre el origen (0,0,0) y la coordenada (R,G,B) en $\mathbb{R}^3$. Véase el gráfico 4.1.

**Gráfico 4.1:** Área cromática en sistema de color RGB.




**Fuente:** www.pce-iberica.es/medidor-detalles-tecnicos/images/colorimetro-pce-rgb-imagen1.jpg

Dada la forma de medir cada una de las características de interés, el número de observaciones a lo largo del experimento es diferente para cada una de las variables.

Cada día se cortarán $r$ papayas de cada nivel para realizar los analizar la acidez pH, los sólidos solubles y dureza, por lo que estas tres características contarán con $r$ observaciones a lo largo del experimento, necesitándose un total de $r * t$ papayas por nivel al inicio del experimento. Las mediciones aunque son tomadas de papayas diferentes, serán relacionadas entre sí, por el nivel de repisa vertical en la refrigeradora donde se coloquen las mismas.

Inicialmente, se seleccionarán $r$ papayas de manera aleatoria, las cuales serán las últimas en ser cortadas, y estas serán objeto del análisis de color.

Dado que las papayas son cortadas, para la variable peso se utilizó le método de imputación definido en el Capítulo 3, de tal manera que se cuente con la información de las $r * t$ papayas por nivel para todo instante de tiempo $t$.

## *4.2. Experimentación Piloto*

Con el fin de determinar el número de réplicas que deberá tenerse en cada nivel del factor *película*, se realizó una experimentación piloto para

tener resultados preliminares. Con los datos resultantes y bajo los supuestos teóricos de pérdida de peso, por medio de simulación se procedió a determinar el número de replicas a utilizarse basándonos en la función de poder.

### *4.2.1. Metodología*

Para la experimentación piloto se seleccionaron doce papayas verdes de la especie “hawaiana” (*Carica Papaya L),* pertenecientes a una misma cosecha, previendo iguales características en los frutos. El experimento piloto se realizó en dos laboratorios de la Facultad de Ingeniería Mecánica y Ciencias de la Producción de la ESPOL. En el laboratorio de Investigación y Desarrollo, se procedió a lavarlas y desinfectarlas por un minuto con hipo clorito de sodio al 2.6% y posteriormente se enjuagaron dos veces con agua destilada.

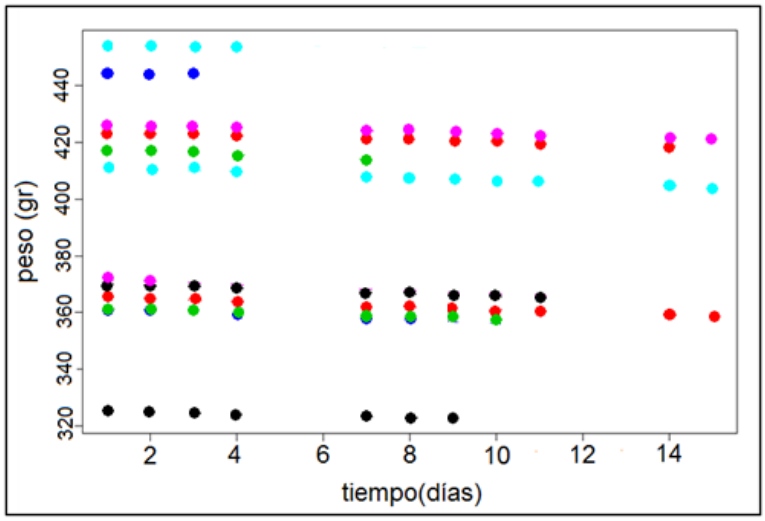
Las papayas fueron puestas en refrigeración de tal manera que se las seleccionó y colocó aleatoriamente, utilizando las tres repisas verticales de la refrigeradora, con el objetivo de confundir el efecto de la posición en la misma. En el laboratorio de Bromatología se procedió pesar las papayas diariamente a excepción de los fines de

semana donde no se tuvo acceso al laboratorio. Una papaya fue cortada a partir del segundo día con el fin de simular el análisis de nivel de acidez, sólidos solubles y dureza.

### *4.2.2. Resultados*

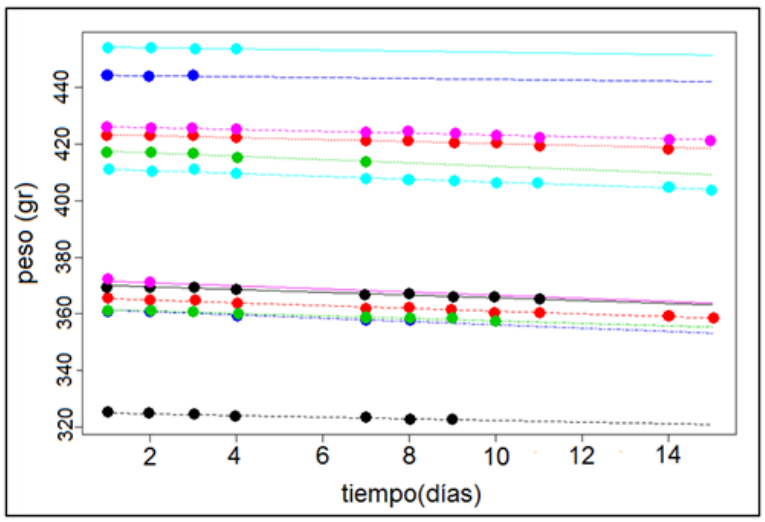
El gráfico 4.2 muestra los pesos en gramos de las doce papayas sometidas al experimento piloto. Los valores resultantes se encuentran en la sección de anexos. Es fácil notar que la perdida se peso se podría suponer lineal, por lo que usar una función de este tipo para representar la pérdida de peso, no sería “descabellado”.

**Gráfico 4.2:** Pesos en gramos de doce papayas medidas en once ocasiones durante los quince días del experimento piloto.



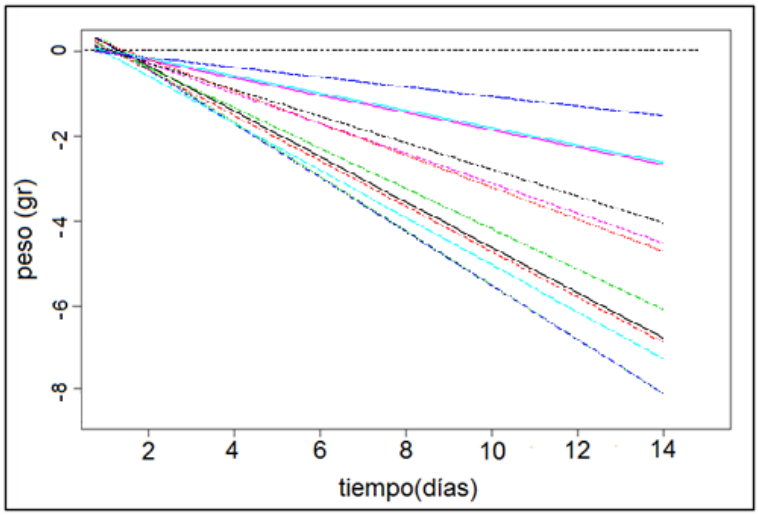
El uso de funciones de mayor grado fue descartado, dado que en los casos donde se tienen pocas observaciones, esto da lugar a tener extrapolaciones erróneas. En el gráfico 4.3, se puede observar las funciones lineales que interpolan las observaciones, las cuales constituyen datos funcionales que fueron hallados por FDA.

**Gráfico 4.3:** Funciones aproximadas de los pesos en gramos de doce papayas medidas en once ocasiones durante los quince días del experimento piloto.




Dado que las papayas tienen diferentes pesos iniciales, fue necesario analizar la variable aleatoria *"pérdida de peso en gramos"* con el propósito de *encerar* las funciones, ya que la pérdida de peso al inicio del experimento será siempre cero. En el gráfico 4.4 se observan las funciones de pérdida de pesos.

**Gráfico 4.4:** Funciones de pérdida de peso en gramos de doce papayas medidas en once ocasiones durante los quince días del experimento piloto.



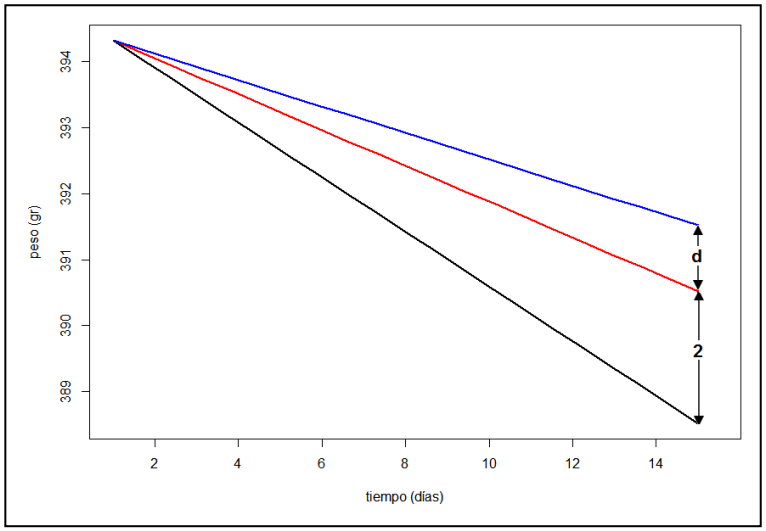

### *4.2.3. Simulación*

Por medio de simulación matemática, con los resultados de la experimentación piloto se simularon 10000 experimentos en las cuales con un nivel de significancia $\alpha$ de 0.05 en cada una de las simulaciones, se realizaron la prueba F de FANOVA y la prueba T de diferencias de medias, esto probando con diferentes tamaños de réplicas y variando los supuestos de los efectos entre niveles. Al final se obtuvieron las Funciones de Poder respectivas para cada prueba.

### *4.2.3.1. Diseño de la Simulación*

Basados en [2], se sabe que el efecto de la aplicación de una película comestible sobre el peso de una papaya al cabo de 15 días es cercano a dos gramos, lo cual fue tomado como supuesto.

**Gráfico 4.5:** Funciones promedio de pérdida de peso en gramos para papayas sin película (negro), con película (rojo) y con película con aceites esenciales (azul).




En el gráfico 4.5 se puede observar el supuesto de dos gramos en el efecto sobre el peso al aplicar una película estándar, mientras que el efecto de aplicar aceites esenciales sobre una película comestible se supone un valor $d$ a variarse, el cual toma los valores 0.50, 0.75, 1.00, 1.25 y 1.50 gramos.

Los diferentes tamaños de replicas a probarse son 3, 4, 5, 10, 15 y 20 papayas, además se supuso que cada uno de los niveles del tratamiento tienen igual varianza.

Dado el supuesto de igualdad de varianzas para la prueba T se utilizará el estimador insesgado de la varianza poblacional $s_p^2(t)$, el cual es la ponderación de las funciones varianzas de los tres niveles, esto es

$$s_p^2(t) = \frac{(n_1 - 1)s_1^2(t) + (n_2 - 1)s_2^2(t) + (n_3 - 1)s_3^2(t)}{n_1 + n_2 + n_3 - 3}. \quad (4.2)$$

Como para este caso el número de réplicas es el mismo para cada nivel, el estimador $s_p^2(t)$ puede ser escrito de manera simplificada como

$$s_p^2(t) = \frac{s_1^2(t) + s_2^2(t) + s_3^2(t)}{3}, \quad (4.3)$$

notando que el estimador de la varianza poblacional coincide con el promedio de las funciones de varianza correspondientes a los tratamientos. El estadístico T se puede expresar de igual manera como

$$T(t) = \frac{\bar{x}_1(t) - \bar{x}_2(t)}{s_p(t)\sqrt{\frac{2}{n}}}. \quad (4.4)$$

Para “generar” las papayas se utilizó la función media de cada

grupo del tratamiento y se les sumo una variación en términos de su desviación estándar. Cada papaya fue generada como

$$Y_{rp}(t) = \hat{\mu}_p(t) + Z\, s_p(t), \qquad (4.5)$$

donde $Y_{rp}(t)$ representa la función de pérdida de peso de la papaya $r$ con tipo de película $p$, $\hat{\mu}_p(t)$ es la función promedio de pérdida de peso para papayas con tipo de película $p$, $Z$ es una variable aleatoria Normal estándar y $s_p(t)$ el estimador insesgado de la función de desviación estándar poblacional.

Una vez que se tuvieron $r$ papayas por nivel, para cada una de las 10000 simulaciones, se realizaron las pruebas funcionales F de FANOVA y la prueba T de diferencias de medias, cuyos estadísticos de prueba se encuentran definidos en (3.13) y (4.4) respectivamente.

La función de poder para la prueba F indica cual es la probabilidad de detectar diferencias significativas en el peso entre los tres niveles del tratamiento, esto es, de papayas sin películas, papayas con películas estándar y con películas con aceites esenciales. De igual manera, la función de poder de la prueba T, nos permitirá calcular la probabilidad de detectar diferencias significativas en el peso al aplicar películas estándar y con aceites esenciales.

#### *4.2.3.2. Algoritmo de Simulación*

La programación del algoritmo de simulación se realizó en el software estadístico R. El código en R de la función que realiza las simulaciones se muestra en la tabla 4.

**Tabla 4:** Algoritmo de simulación en R para determinar el número de réplicas a utilizarse en el experimento.

```
potencia=function(r,d)
{
  base.lineal=create.bspline.basis(range(tiempo),
        norder=2)
  dif1=Data2fd(range(tiempo),c(0,2),base.lineal)
  dif2=Data2fd(range(tiempo),c(0,d),base.lineal)
  m0=media
  m1=m0+dif1
  m2=m1+dif2
  s=as.vector(eval.fd(c(1:15),de))
  numrechF=0
  numrechT=0
  nt=length(tiempo)
  Fcrit=qf(alpha=0.05,3-1,r*3-3,lower.tail=FALSE)
  Tcrit=qt(alpha=0.05,r*3-3,lower.tail=FALSE)
  tratamiento<<-factor(rep(0:2,rep(r,3)))
  for(i in 1:10000)
  {
    muestra0=vector(mode="list",length=r)
    for(j in 1:r)
    {
      muestra0[[j]]=m0+(de*rnorm(1))
    }
    muestra1=vector(mode="list",length=r)
    for(j in 1:r)
    {
      muestra1[[j]]=m1+(de*rnorm(1))
    }
    muestra2=vector(mode="list",length=r)
    for(j in 1:r)
    {
      muestra2[[j]]=m2+(de*rnorm(1))
    }
```

**Tabla 4 (Continua):** Algoritmo de simulación en R para determinar el número de replicas a utilizarse en el experimento.

```
    muestrav0=sapply(muestra0,function(x)
          as.vector(eval.fd(evalarg=c(1:15),x)))
    muestrav1=sapply(muestra1,function(x)
          as.vector(eval.fd(evalarg=c(1:15),x)))
    muestrav2=sapply(muestra2,function(x)
          as.vector(eval.fd(evalarg=c(1:15),x)))
    muestrav0=apply(muestrav0,2,function(x) x-x[1])
    muestrav1=apply(muestrav1,2,function(x) x-x[1])
    muestrav2=apply(muestrav2,2,function(x) x-x[1])
    muestra=cbind(muestrav0,muestrav1,muestrav2)
    muestrafd<<-Data2fd(c(1:15),muestra,base.lineal)

    #PRUEBA T
    media1=apply(muestrav1,1,mean)
    media2=apply(muestrav2,1,mean)
    var0=apply(muestrav0,1,var)
    var1=apply(muestrav1,1,var)
    var2=apply(muestrav2,1,var)
    const=sqrt(2/r)
    sp2=(var0+var1+var2)/3
    sp=sqrt(sp2)
    t=(media2-media1)/(sp*const)

    #PRUEBA F
    reg=fRegress(muestrafd~tratamiento)
    yvec = eval.fd(tfino, reg$yfdPar$fd)
    yhatvec = eval.fd(tfino, reg$yhatfdobj$fd)
    SCE=apply((yvec-yhatvec)^2,1,sum)
    SCT=(apply(yvec,1,var))* (r*3-1)
    SCR=SCT-SCE
    MCE=SCT/(r*3-r)
    MCR=SCR/(r-1)
    F=MCR/MCE
  }
    #FUNCION DE PODER
    {
      #PRUEBA T
      maxt=max(t,na.rm=TRUE)
      if(maxt>Tcrit) numrechT=numrechT+1

      # PRUEBA F
      maxf=max(f$F)
      if(maxf>Fcrit) numrechF=numrechF+1
    }
      c(numrechF/10000,numrechT/10000)
}
```

### *4.2.3.3. Resultados de la simulación*

En la tabla 5 se observan los resultados de la simulación. Se muestran los distintos valores que toma la Función de Poder para las pruebas F y T, para los diferentes tamaños de replicas y las diferentes distancias.

**Tabla 5:** Funciones de Poder de las pruebas F y T realizadas por medio de simulación, para diferentes tamaños de réplicas y distancias.

| | | d = 0.50 | d = 0.75 | d = 1.00 | d = 1.25 | d = 1.50 |
|---|---|---|---|---|---|---|
| r = 3 | F | 0.9377 | 0.9675 | 0.9856 | 0.9958 | 0.9992 |
| | t | 0.2533 | 0.4323 | 0.6238 | 0.7907 | 0.9002 |
| r = 4 | F | 0.9576 | 0.9827 | 0.9946 | 0.9987 | 0.9996 |
| | t | 0.3275 | 0.5484 | 0.7656 | 0.9055 | 0.9704 |
| r = 5 | F | 0.9921 | 0.9987 | 0.9999 | 1.0000 | 1.0000 |
| | t | 0.3870 | 0.6470 | 0.8557 | 0.9586 | 0.9932 |
| r = 10 | F | 0.9972 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | t | 0.6381 | 0.9075 | 0.9882 | 0.9998 | 1.0000 |
| r = 15 | F | 0.9997 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | t | 0.7809 | 0.9793 | 0.9995 | 1.0000 | 1.0000 |
| r = 20 | F | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | t | 0.8864 | 0.9972 | 1.0000 | 1.0000 | 1.0000 |

Con los resultados de las simulaciones se decidió utilizar tres replicas por cada nivel, ya que asumiendo una diferencia promedio de entre 1 gramo y 1.25 gramos entre papayas con películas estándar y películas con aceites esenciales, se obtiene un poder

de 0.62 a 0.80 para la prueba T y aproximadamente 0.99 para la prueba F. Utilizar un número de réplicas pequeño, facilita de gran manera la experimentación dadas las restricciones de espacio y uso de los equipos en los laboratorios.

**Gráfico 4.6:** Funciones de Poder para la prueba F para los diferentes tamaños de réplica y distancia a probarse.

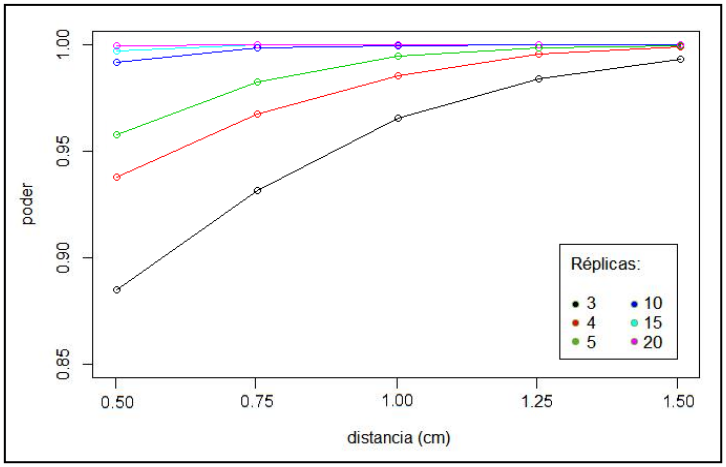
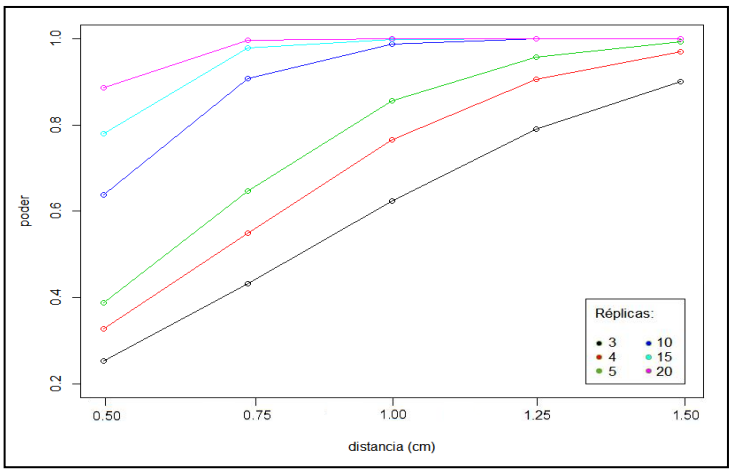



**Gráfico 4.7:** Funciones de Poder para la prueba T para los diferentes tamaños de réplica y distancia a probarse.



## *4.3. Diseño del Experimento*

Una vez determinado el tamaño de muestra se sabe que serán tres, las papayas que serán cortadas día a día para los análisis invasivos por lo que el número de réplicas son: 30 papayas por nivel para el análisis de peso y 3 papayas por nivel para el análisis de acidez, sólidos solubles, dureza y color.

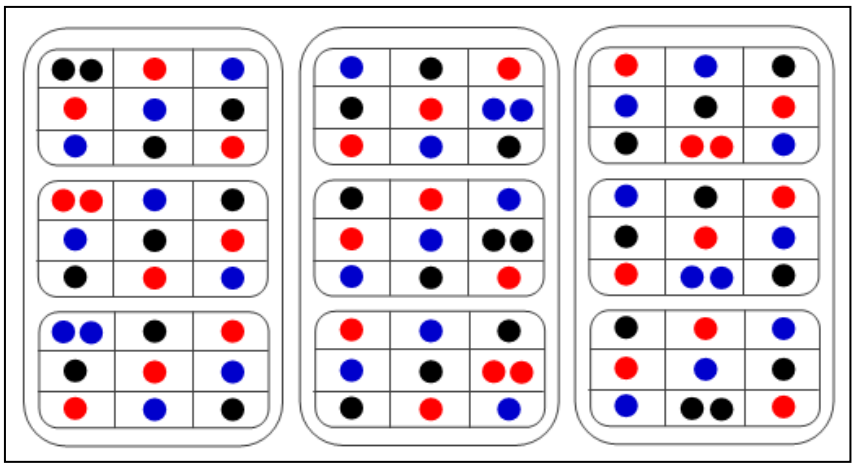
### *4.3.1. Confusión*

Teniendo inicialmente 90 papayas en total, estas deberán ser distribuidas en tres refrigeradoras de iguales características, en las que se usaron sus tres niveles de repisas verticales; es por esto, que es de preocupación los posibles efectos que pudieran tener el uso de distintas refrigeradoras, el nivel vertical de repisa y la posición de los frutos dentro de ellas.

Es por esto, que elaboramos un diseño en el que las papayas de cada nivel deben ser colocadas de tal manera que estos efectos no se confundan con el efecto de película. Para esto, definiremos los bloques *"refrigerador"*, *"nivel de repisa vertical del refrigerador"* y

*“posición de la papaya en la repisa”.* El diseño propuesto con el fin de confundir los efectos anteriormente mencionados se muestra en el gráfico 4.8.

**Gráfico 4.8:** Posición espacial de las papayas en las refrigeradoras durante el experimento. Obsérvese las papayas sin película (negro), con película (rojo) y con película con aceites esenciales (azul).



En el gráfico 4.8, se puede observar que existe una única posición por repisa donde se encuentran dos papayas; una de ellas va a ser cortada el primer día del experimento, mientras que la restante va a ser objeto del análisis del color durante todo el experimento. La selección del grupo de papayas a cortarse por día se realizó de manera aleatoria la cual se muestra paso a paso en la sección de anexos.

# CAPÍTULO V

## 5. RESULTADOS DEL EXPERIMENTO

### *5.1. Introducción*

En este capítulo se presenta el análisis univariado de las variables de interés, un breve análisis bivariado, así como los resultados obtenidos del Análisis de Varianza Funcional y Contrastes de Hipótesis de interés aplicados al modelo estadístico del Diseño Experimental en estudio. Tal como se mencionó en capítulos anteriores, todos los resultados obtenidos han sido realizados en el software estadístico libre R.

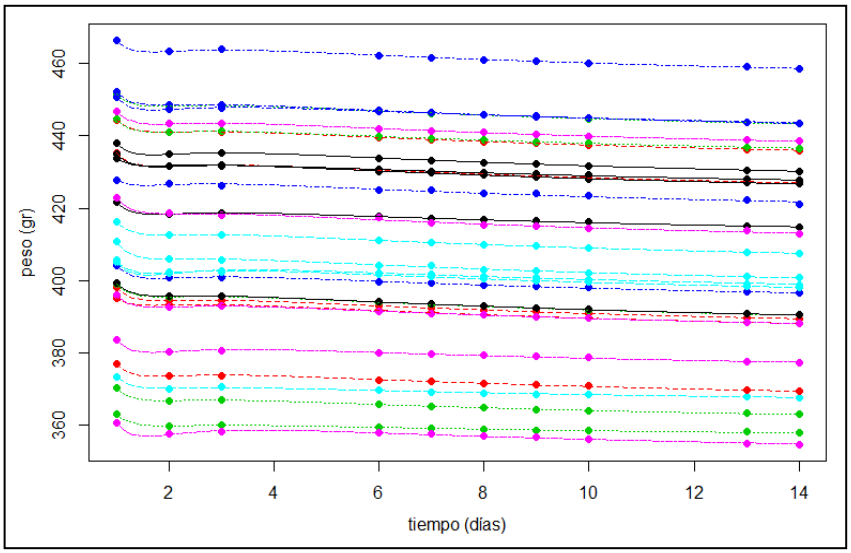
### *5.2. Ajuste de Datos*

Dada la naturaleza de cada una de las variables analizadas, cada una de ellas fue modelizada por funciones con distintos sistemas bases. Las variables nivel de pH, sólidos solubles, dureza y color utilizaron un sistema B-spline mientras que para representar las funciones de pérdida de peso fue necesario utilizar una sistema base exponencial.

**Tabla 6:** Sistemas bases seleccionados para modelizar las variables de interés.

| Variable | Sistema Base | Parámetros |
|---|---|---|
| Peso | Exponencial | $\{1, e^{-0.1}, e^{-1}, e^{-3}\}$ $\lambda = 10^{-4}$ |
| Nivel de pH | B-Spline | $\text{nbasis} = 6, \quad \text{norder} = 4$ |
| Sólidos Solubles | B-Spline | $\text{nbasis} = 7, \quad \text{norder} = 4$ |
| Dureza | B-Spline | $\text{nbasis} = 7, \quad \text{norder} = 4$ |
| Color | B-Spline | $\text{nbasis} = 8, \ \text{norder} = 4, \ \lambda = 0.479$ |

**Gráfico 5.1:** Funciones ajustadas para las observaciones correspondientes al peso en gramos para papayas sin películas.




Para obtener un buen ajuste para la variable color fue necesario utilizar penalización por aspereza cuyo parámetro $\lambda$ fue hallado por *Generalized Cross-Validation*, método que fue previamente definido en

el Capítulo 2. De igual manera para la variable peso se utilizó una penalización de aspereza de $\lambda = 10^{-4}$ para un mejor ajuste con las observaciones. Las observaciones faltantes fueron halladas por el método de imputación definido en el Capítulo 3, y los elementos de la base exponencial fueron elegidos bajo el criterio del VIF.

## *5.3. Análisis Univariado*

A continuación se presentarán las funciones análogas a las estadísticas descriptivas más relevantes para cada una de las variables, además de ciertas funciones que sirvan como indicadores de características de importancia. Las papayas sin película, con películas comestibles y con películas comestibles con aceites esenciales se representan con los colores negro, rojo y azul respectivamente.

- ***Peso***

  En el gráfico 5.2 se pueden observar las 90 funciones correspondientes a las pérdidas de peso (gramos) de las papayas a lo largo del experimento. En promedio, las papayas que no poseen algún tipo de película comestible son las que pierden más peso,

siendo evidente que casi la mitad del peso total perdido a lo largo de los quince días del experimento, se pierde en los primeros dos días, siendo el tercer día donde la papaya aumenta su peso levemente antes de caer lentamente.

**Gráfico 5.2:** Funciones de pérdida de peso en gramos.

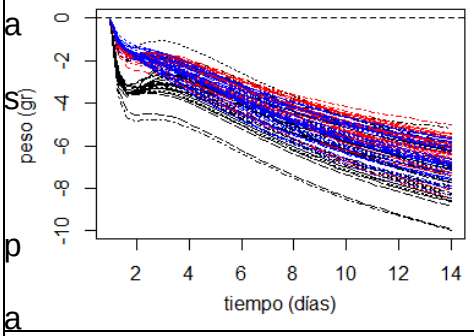



**Gráfico 5.3:** Funciones promedio de pérdida de peso en gramos.

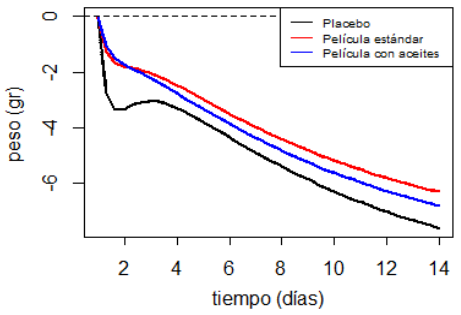



**Gráfico 5.4:** Funciones desviaciones estándar de pérdida de peso en gramos.

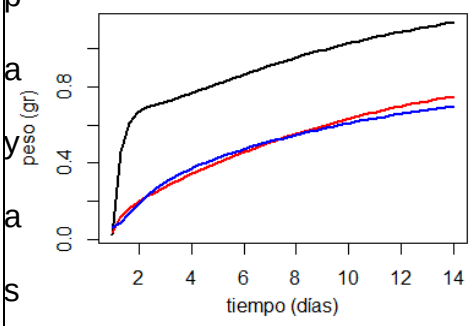
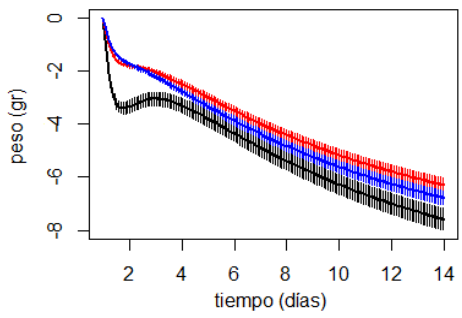



**Gráfico 5.5:** Intervalos de Confianza del 95% para la funciones medias poblacionales.




Las papayas a las cuales les fue aplicadas las películas comestibles estándares, fueron las que en promedio tuvieron la menor pérdida de peso casi durante toda la duración del experimento, donde al día final se obtuvo una diferencia de más de un gramo con respecto al control, y aproximadamente medio gramo con las papayas con películas

comestibles con aceites esenciales. La función desviación estándar para las papayas de control, es mucho mayor que las funciones desviaciones estándar de las papayas con películas comestibles, siendo estas dos últimas funciones casi iguales. Por tanto, la aplicación de películas comestibles no solo reduce la pérdida de peso, sino que reduce la variabilidad de la misma, por lo que facilitaría el realizar predicciones. Con un 95% de confianza se puede asegurar que el promedio de pérdida de peso para papayas sin películas comestibles siempre será mayor que si se aplicará algún tipo de película.

- ***Nivel de pH***

Durante el experimento como era de esperarse, el potencial de Hidrógeno pH aumentó, dado que la fruta al madurar pierde su acidez y se torna alcalina. En promedio, la aplicación de la película comestible con aceites esenciales acelera la maduración de la fruta (en términos de su acidez) teniendo los mayores niveles de pH en el experimento hasta el día 10 donde son superadas levemente por las papayas sin películas las cuales se mantuvieron ácidas en mayor manera hasta aproximadamente el día 6 y luego perdieron acidez rápidamente (ver gráfico 5.9). Al final de los quince días del

experimento, fueron las papayas con películas comestibles estándar quienes se mantuvieron más ácidas.

**Gráfico 5.6:** Funciones de nivel de acidez pH para nueve observaciones durante la realización del experimento.

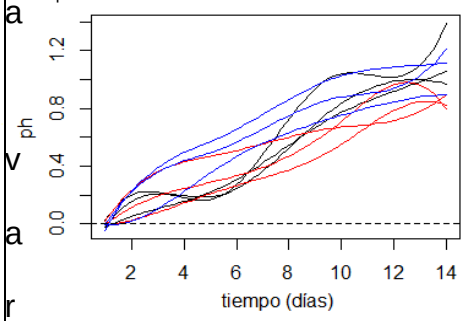



**Gráfico 5.7:** Funciones promedio de aumento del nivel pH durante la realización del experimento.

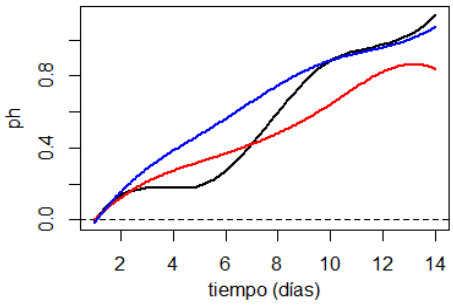



**Gráfico 5.8:** Funciones desviaciones estándar del aumento del nivel pH durante la realización del experimento.

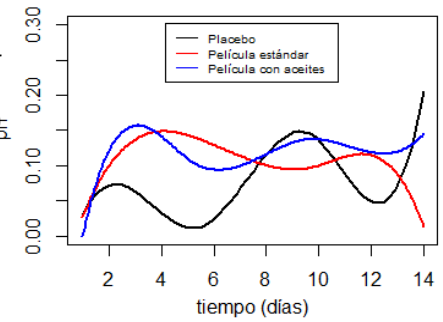
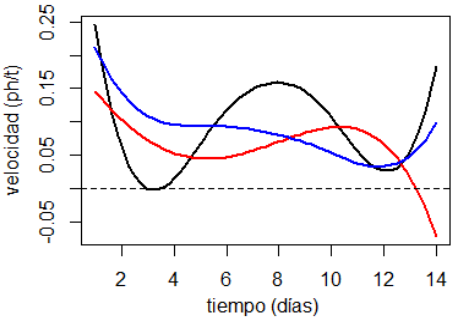



**Gráfico 5.9:** Funciones de velocidad promedio del aumento del nivel pH durante la realización del experimento.




Variabilidad

La variabilidad del aumento del nivel de acidez pH es casi impredecible y parecería verse afectado por factores no contemplados en este experimento. Al menos durante los primeros siete días parecería que las papayas de control son las que poseen menor variabilidad la cual aumenta en los días finales.

- ***Sólidos solubles***

En promedio, una papaya sin algún tipo de película comestible aumenta 4º brix durante la duración del experimento, mas si se aplicase algún tipo de película comestible, es decir con o sin aceites esenciales, la cantidad de sólidos solubles al final del experimento se va a ver reducido a aproximadamente la mitad.

**Gráfico 5.10:** Funciones de aumento de sólidos solubles para nueve observaciones durante el experimento.

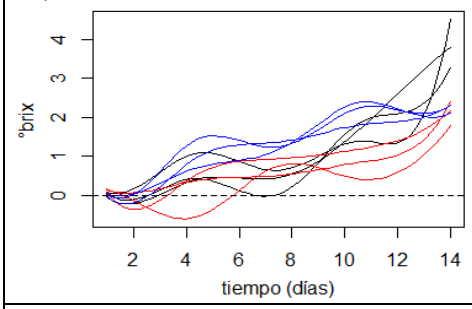



**Gráfico 5.11:** Funciones promedio de aumento de sólidos solubles para cada nivel durante el experimento.

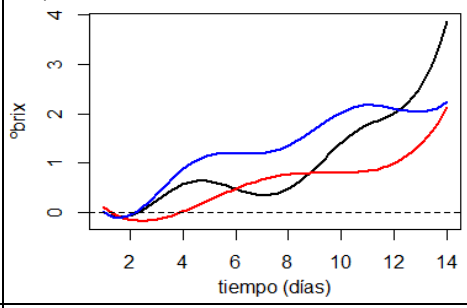



**Gráfico 5.12:** Funciones desviación estándar del aumento de sólidos solubles para cada nivel durante el experimento.

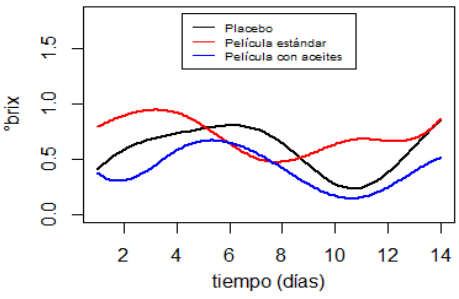
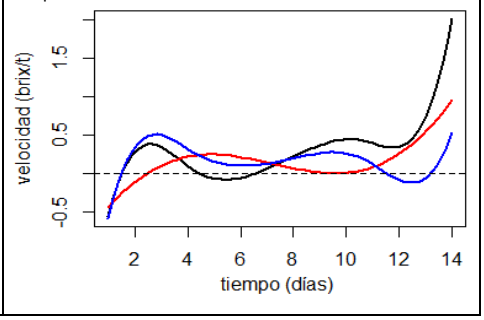



**Gráfico 5.13:** Funciones de velocidad del aumento de sólidos solubles para cada nivel durante el experimento.




El cambio en el nivel de sólidos solubles para las papayas sin películas se mantuvo por debajo de una unidad absoluta hasta el

noveno día del experimento, donde a partir de este momento la velocidad del aumento se dispara exponencialmente. Aunque la aplicación de una película comestible estándar fue la que dio mejores resultados para el aumento promedio de sólidos solubles, esta es la que contiene mayor variabilidad.

- ***Dureza***

Como era de esperarse la dureza es inversamente proporcional al tiempo transcurrido durante la experimentación, mas cabe destacar que cuando no se aplican películas comestibles en las papayas, en promedio, se obtuvo la mayor dureza en el fruto durante en todo el experimento.

Al aplicarse películas comestibles estándares durante la experimentación se observó menor dureza en el fruto aunque al final del experimento se hayan obtenido iguales resultados a que si no se aplicara.

La aplicación de películas comestibles con aceites esenciales parecería tener efectos negativos al tenerse los menores registros de durezas en papayas durante toda la realización del experimento.

**Gráfico 5.14:** Funciones de pérdida de dureza en **papayas** durante el experimento.

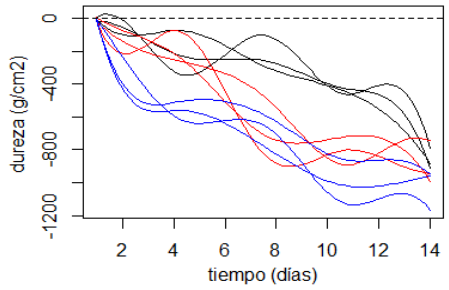



Gráfico **5.15:** Funciones promedio de pérdida de dureza para cada nivel durante el experimento.

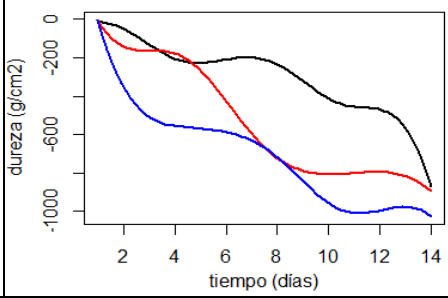



**Gráfico 5.16:** Funciones desviación estándar de la pérdida de dureza para cada nivel durante el experimento.

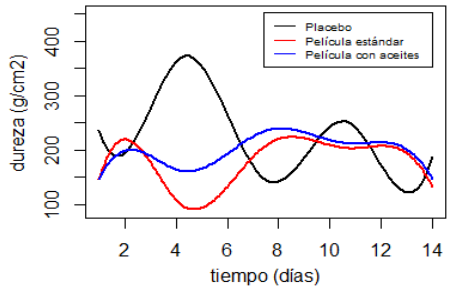
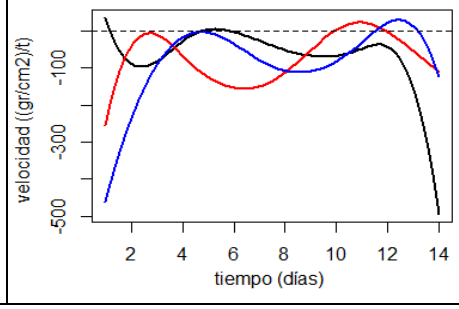



**Gráfico 5.17:** Funciones de velocidad de la pérdida de dureza para cada nivel durante el experimento.




Aunque al final del día catorce se obtuvieron iguales resultados en dureza entre el control y las papayas con películas comestibles estándar, si se deseara predecir la pérdida de dureza en los próximos días al experimento, se tendría que las papayas sin películas perderían dureza más rápidamente dado que su velocidad de pérdida de dureza al final de experimento es aproximadamente tres veces mayor a la velocidad de pérdida de peso para las papayas con películas comestibles estándares (ver gráfico 5.17).

Durante los primeros días del experimento, las papayas sin películas comestibles fueron las que tuvieron mayor variabilidad,

mientras que para las papayas con películas comestibles de ambos tipos su variabilidad luce algo similar durante el experimento.

- ***Color***

El uso de películas comestibles afecta de manera positiva a disminuir los efectos de la maduración de la fruta sobre la coloración de la misma, siendo aún más efectivo, el uso de películas comestibles estándares. Es  partir del sexto día donde se puede evidenciar diferencias en la coloración promedio de las papayas para cada uno de los respectivos grupos.

La variabilidad para las papayas que les fue aplicadas películas comestibles estándar fue mayor para los primeros días 10 días del experimento, mientras que las papayas de control fueron quienes presentaron menor variabilidad.

En los gráficos 5.18, 5.19 y 5.20 se pueden observar una representación gráfica de las tonalidades reales que tomaron los frutos a lo largo del experimento, las papayas de control, con películas comestibles estándar y películas comestibles con aceites esenciales respectivamente.

**Gráfico 5.18:** Representación gráfica de la coloración de las papayas sin películas comestibles durante el experimento.

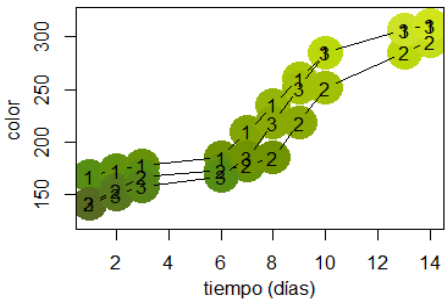



**Gráfico 5.19:** Representación gráfica de la coloración de las papayas con películas comestibles estándar durante el experimento.

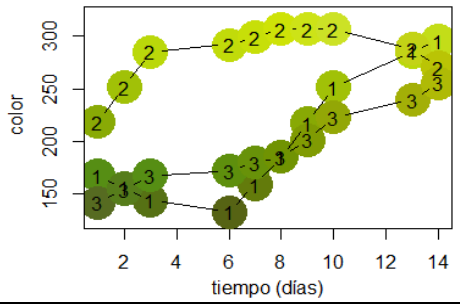



**Gráfico 5.20:** Representación gráfica de la coloración de las papayas con películas comestibles con aceites esenciales durante el experimento.

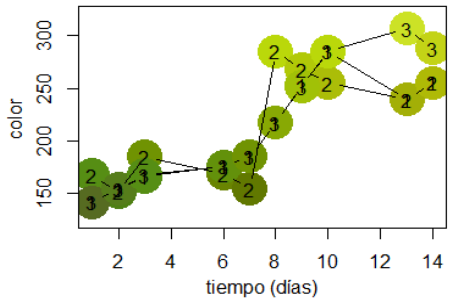



**Gráfico 5.21:** Funciones de cambio de color en papayas durante la realización del experimento.

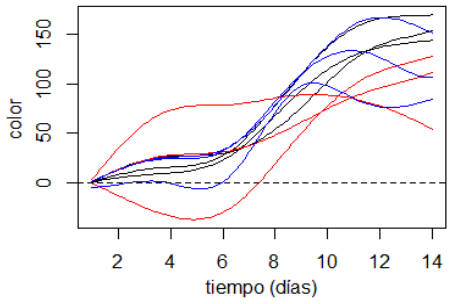



**Gráfico 5.22:** Funciones promedio del cambio de color en papayas para cada nivel durante la realización del experimento.

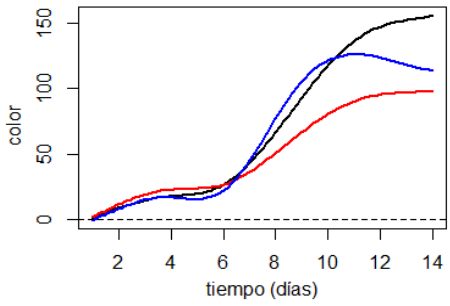
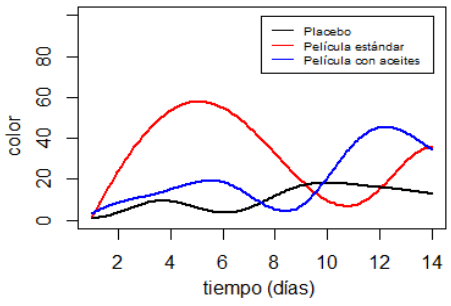



**Gráfico 5.23:** Funciones desviación estándar del cambio de color en papayas para cada nivel durante la realización del experimento.

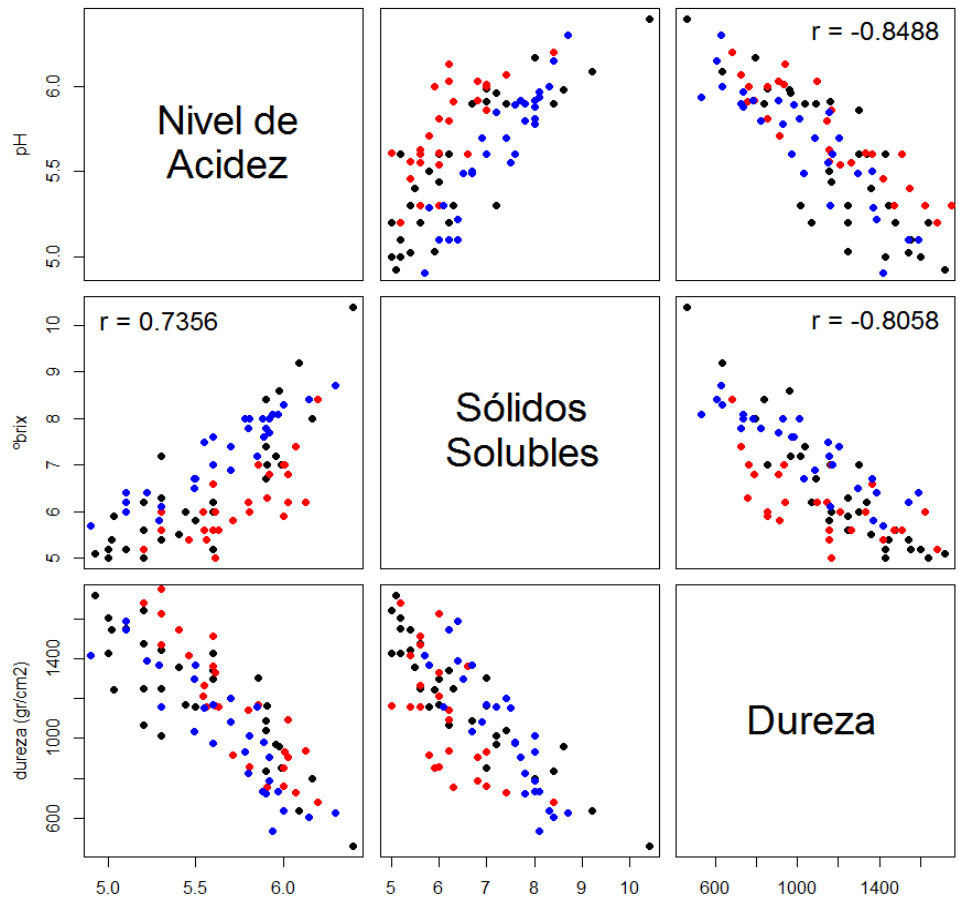



## *5.4. Análisis Bivariado*

En el gráfico 5.24 se puede observar una Matriz de Diagrama de Dispersión en la cual se representan de manera simultánea los

diagramas de dispersión de las variables de interés acidez, sólidos solubles y dureza. Fue de interés la relación entre estas variables pues para ellas se utilizaron las mismas unidades experimentales en cada instante de tiempo $t$.

**Gráfico 5.24:** Matriz de Diagrama de Dispersión.




Las variables nivel de pH, sólidos solubles y dureza, se encuentran fuertemente relacionadas entre sí ya que estas crecen o disminuyen

casi de manera monótona durante el experimento. De este análisis podemos decir que:

- El nivel de pH y los sólidos solubles se encuentran correlacionados de manera positiva, pues a medida que aumenta el nivel pH del fruto y se torna menos ácido, los sólidos solubles aumentarán dando una mayor dulzura al fruto.

- El nivel de pH y la dureza son el par de variables que se encuentran más fuertemente correlacionados ($r = -0.8488$). Su relación es negativa pues a medida que el fruto pierde su acidez también este tiende a hacerse más blando.

- Los sólidos solubles y la dureza se encuentran correlacionados negativamente pues a medida que el fruto se torne más dulce por el proceso de maduración este se tornará más blando.

## *5.5. Análisis de Varianza Funcional*

En el análisis de varianza funcional y el análisis de la prueba T para diferencia de medias se presenta a continuación para cada una de

las variables de interés. En los gráficos correspondientes a las funciones análogas a los estadísticos de prueba $F(t)$ y $T(t)$ las líneas horizontales segmentadas representan los percentiles $F_{0.90}$ y $F_{0.95}$ para la prueba F, y los percentiles $T_{0.05}$ y $T_{0.10}$ para la prueba T, esto con el fin de delimitar la zona de rechazo, incertidumbre y no rechazo; de igual manera en los gráficos de la función $P_{VALUE}$ se ha delimitado rectas horizontales en los valores 0.05, 0.10, esto con el mismo fin.

- ***Peso***

  La Potencia de Explicación del Modelo fue alta únicamente durante los tres primeros días del experimento donde luego disminuyó, manteniéndose casi constante alrededor de un valor de 30%. Obsérvese la función $\mathrm{RSQ}(t)$ en el gráfico 5.26.

  La baja potencia de explicación se debe al crecimiento de la función de la Suma Cuadrática del Error, por lo que una gran parte de la variabilidad de la variable de respuesta no es explicada por el factor "tipo de película" sino por otros factores no tomados en cuenta en este experimento o simplemente factores no controlables. Véase gráfico 5.25. La función análoga al estadístico de prueba F tomó valores exorbitantes y como consecuencia el valor p de la prueba

siempre fue cero durante la realización del experimento, es por esto que al rechazar la hipótesis nula concluimos que siempre existen diferencias significativas sobre el peso de la papaya, al aplicar algún tipo de película comestible, mas sin saber cuál de ellas dos es la más efectiva.

**Gráfico 5.25:** Funciones correspondientes a las Sumas Cuadráticas de Regresión, Error y Total.

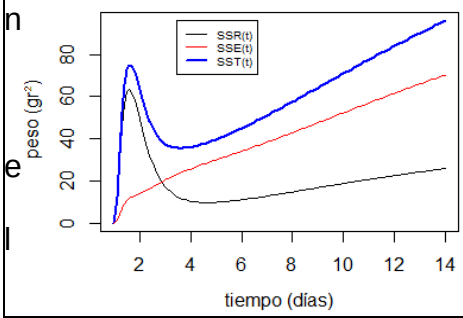



**Gráfico 5.26:** Función $\mathrm{RSQ}(t)$ análoga al Coeficiente de Determinación del Modelo $R^2$.

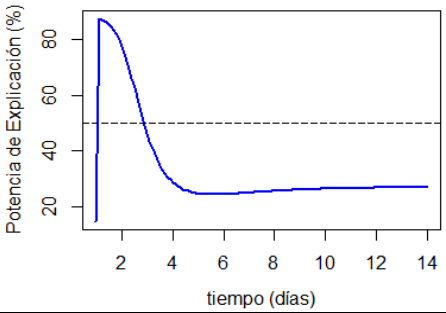



**Gráfico 5.27:** Función $F(t)$ análoga al estadístico de prueba F para el Análisis de Varianza.

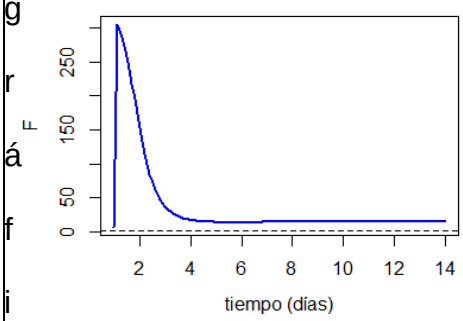



**Gráfico 5.28:** Función $P_{VALUE}$ análoga al valor p o nivel de significancia de la prueba F para todo $t$.

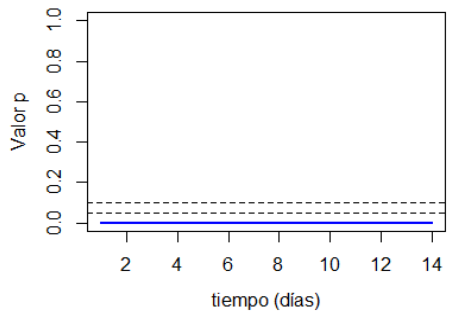
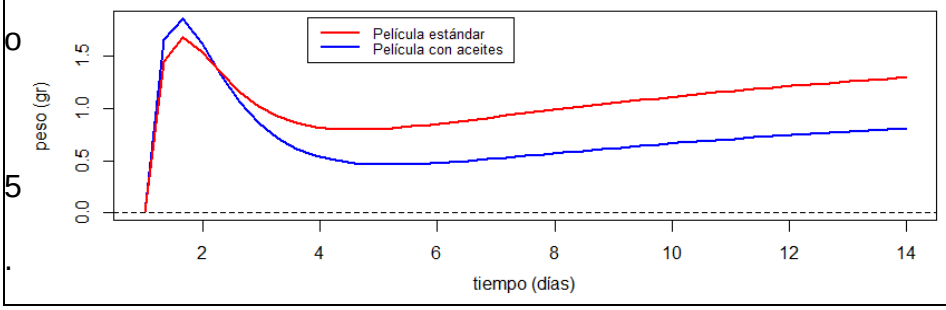



**Gráfico 5.29:** Funciones de efectos $\widehat{\alpha_1}$ y $\widehat{\alpha_2}$ sobre el peso de papayas sin películas (control), al aplicarse películas comestibles estándar y películas comestibles con aceites esenciales respectivamente. Es deseable que el efecto sea positivo.



En 5.29 podemos corroborar la información anterior observando que las funciones correspondientes a los efectos de la aplicación de las películas comestibles son positivas y no cercanas a cero por estos efectos podrían considerarse significativos. El efecto de la aplicación la película con aceites esenciales resulta mayor únicamente durante el primer día del experimento; como resultado se rechazó $H_o$ para la prueba T de diferencias de medias durante este período de tiempo. Véase gráfico 5.31.

**Gráfico 5.30:** Función $T(t)$ análoga al estadístico de prueba T para diferencia de medias. Las varianzas se supusieron iguales.

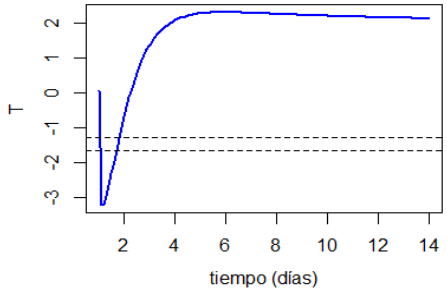
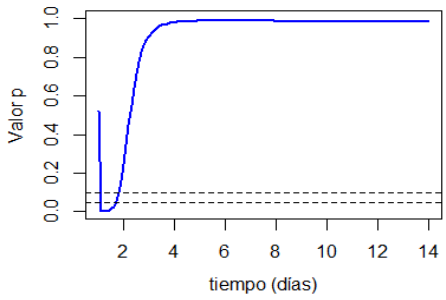



**Gráfico 5.31:** Función $P_{VALUE}$ análoga al valor p o nivel de significancia de la prueba T para todo $t$.




- ***Nivel de pH***

La Potencia de Explicación del Modelo fue medianamente alta a partir del cuarto día del experimento casi siempre manteniéndose por encima del 50% de explicación. Véase gráfico 5.32. Alrededor del sexto día del experimento la Potencia de Explicación del Modelo

alcanzó un 71.86% donde el valor p se encontró en zona de rechazo, dejando en evidencia que para este período de tiempo comprendido ente el quinto y séptimo día existen diferencias significativas en la acidez del fruto al aplicar películas comestibles de cualquier tipo.

**Gráfico 5.32:** Funciones correspondientes a las Sumas Cuadráticas de Regresión, Error y Total

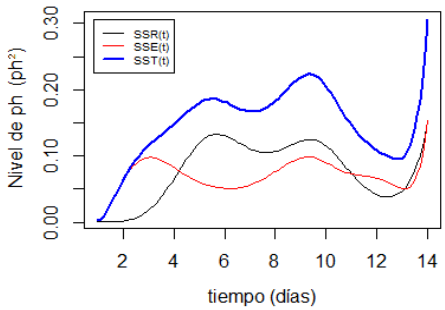



**Gráfico 5.33:** Función $\mathrm{RSQ}(t)$ análoga al Coeficiente de Determinación del Modelo $R^2$.

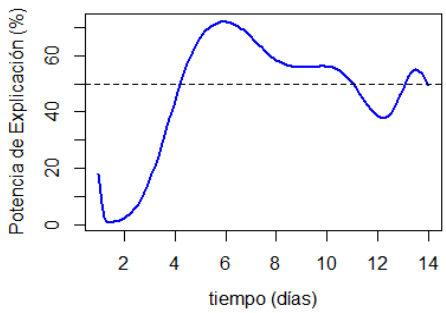



**Gráfico 5.34:** Función $F(t)$ análoga al estadístico de prueba F para el Análisis de Varianza.

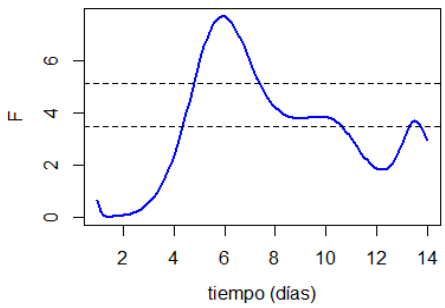



**Gráfico 5.35:** Función $P_{VALUE}$ análoga al valor p o nivel de significancia de la prueba F para todo $t$.

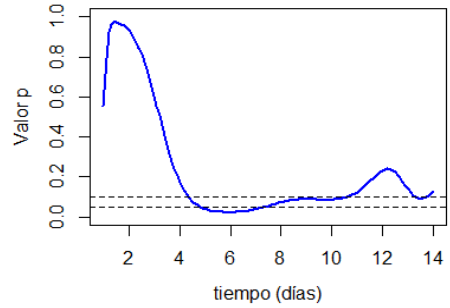
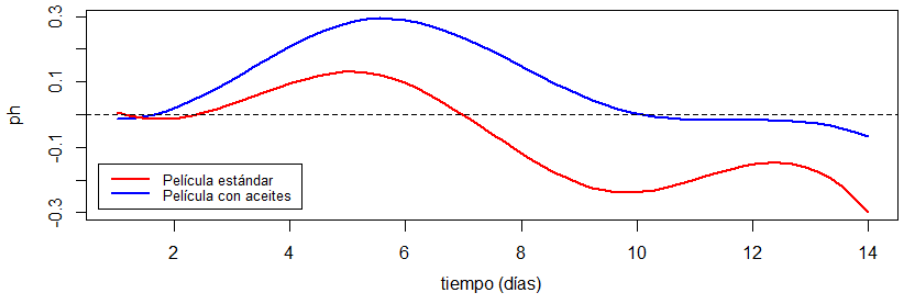



**Gráfico 5.36:** Funciones de efectos $\widehat{\alpha_1}$ y $\widehat{\alpha_2}$ sobre el nivel de pH de papayas sin películas (control), al aplicarse películas comestibles estándar y películas comestibles con aceites esenciales respectivamente. Es deseable que el efecto sea negativo.



En el gráfico 5.36, al observar los efectos al aplicar cada tipo de película comestible se puede observar que el efecto de la aplicación de películas comestibles estándar es casi siempre negativo durante la realización del experimento, siendo esto favorable pues se pretende que el nivel de pH se mantenga bajo con el fin que el fruto se mantenga ácido, siendo esto indicador de la falta de la maduración del fruto.

El efecto de la aplicación de películas comestibles con aceites esenciales es siempre menor que el de las películas comestibles estándar lo cual es evidenciado en el gráfico 5.38 donde el valor p correspondiente a la prueba unilateral de diferencia de funciones medias es siempre mayor a 0.10, por lo que en ningún momento del experimento se puede considerar favorable la aplicación de películas comestibles con aceites esenciales con el fin de preservar el nivel de pH en el fruto.

**Gráfico 5.37:** Función $T(t)$ análoga al estadístico de prueba T para diferencia de medias. Las varianzas se supusieron desiguales.

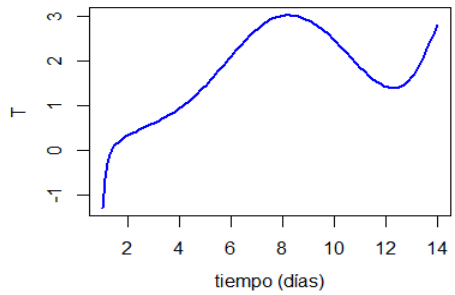
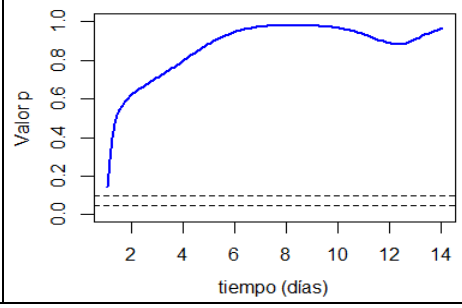



**Gráfico 5.38:** Función $P_{VALUE}$ análoga al valor p o nivel de significancia de la prueba T para todo $t$.



- ***Sólidos solubles***

La Potencia de Explicación del Modelo fue alta a partir del día cuatro del experimento, tomando su valor máximo alrededor del día noveno donde la Potencia de Explicación del modelo fue 91.79%.

**Gráfico 5.39:** Funciones correspondientes a las Sumas Cuadráticas de Regresión, Error y Total

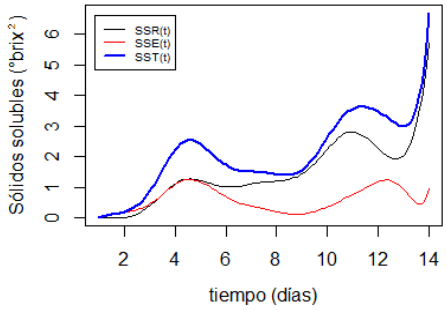



**Gráfico 5.40:** Función $\mathrm{RSQ}(t)$ análoga al Coeficiente de Determinación del Modelo $R^2$.

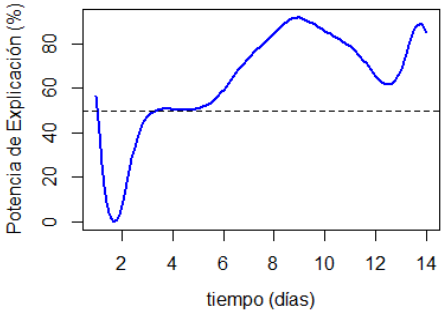



**Gráfico 5.41:** Función $F(t)$ análoga al estadístico de prueba F para el Análisis de Varianza.

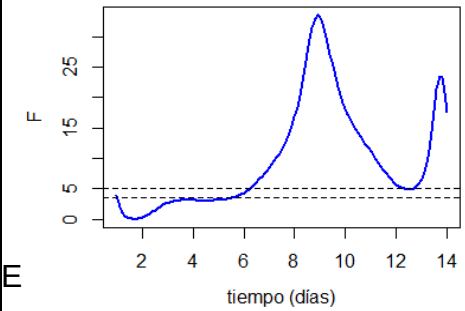
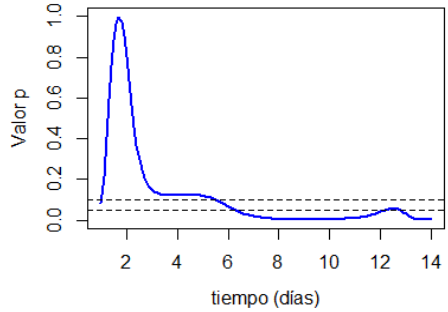
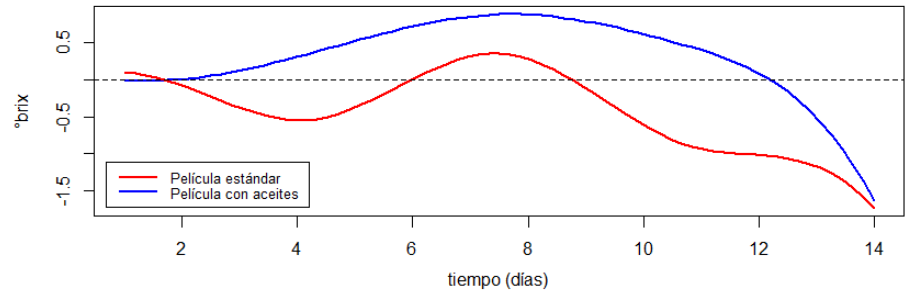



**Gráfico 5.42:** Función $P_{VALUE}$ análoga al valor p o nivel de significancia de la prueba F para todo $t$.




E

**Gráfico 5.43:** Funciones de efectos $\widehat{\alpha_1}$ y $\widehat{\alpha_2}$ sobre los sólidos solubles de papayas sin películas (control), al aplicarse películas comestibles estándar y películas comestibles con aceites esenciales respectivamente. Es deseable que el efecto sea negativo.



A partir del sexto día del experimento el valor p correspondiente a la prueba F de diferencias significativas es menor que 0.05 por lo que se rechaza $H_0$ que postula que no existen diferencias significativas en la aplicación de películas comestibles en las papayas, por lo que se concluye que sí existen diferencias significativas en la aplicación.

En el análisis individual de los efectos sobre los sólidos solubles al aplicar los dos diferentes tipos de películas comestibles se puede observar que el efecto correspondiente a la aplicación de películas comestibles con aceites esenciales es positivo para la mayoría del tiempo, lo cual no es favorable ya que produciría más sólidos solubles que si no se aplicase película alguna. Véase gráfico 5.43.

El análisis de efectos individuales es corroborado nuevamente al evidenciarse en 5.45 que jamás se rechaza $H_0$ que postula que el efecto de la aplicación de aceites esenciales en la película es nulo.

**Gráfico 5.44:** Función $T(t)$ análoga al estadístico de prueba T para diferencia de medias. Las varianzas se supusieron desiguales.

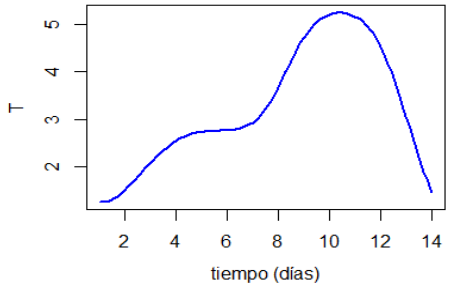
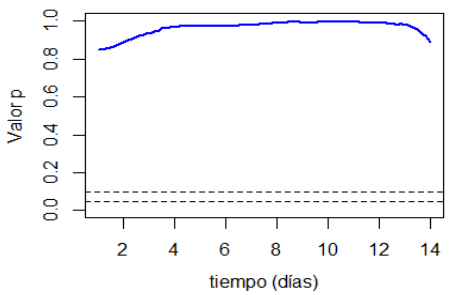



**Gráfico 5.45:** Función $P_{VALUE}$ análoga al valor p o nivel de significancia de la prueba T para todo $t$.



- ***Dureza***

Con respecto a la variable dureza, la Suma Cuadrática del Error fue siempre baja dando como consecuencia que para toda la realización del experimento la Potencia de Explicación sea siempre alta tomando incluso valores mayores al 80%. Véase gráfico 5.47.

**Gráfico 5.46:** Funciones correspondientes a las Sumas Cuadráticas de Regresión, Error y Total

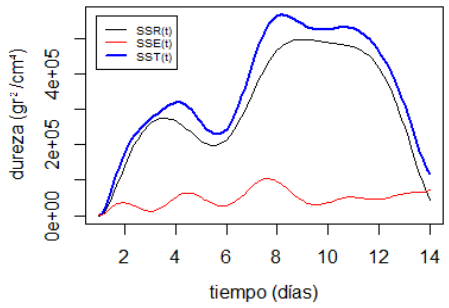



**Gráfico 5.47:** Función $\mathrm{RSQ}(t)$ análoga al Coeficiente de Determinación del Modelo $R^2$.

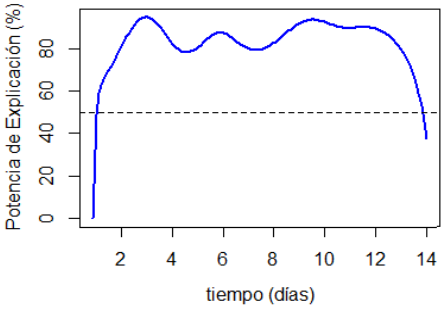



**Gráfico 5.48:** Función $F(t)$ análoga al estadístico de prueba F para el Análisis de Varianza.

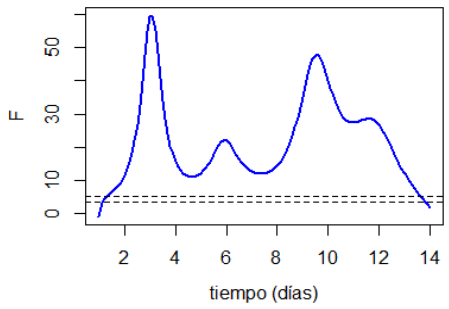



**Gráfico 5.49:** Función $P_{VALUE}$ análoga al valor p o nivel de significancia de la prueba F para todo $t$.

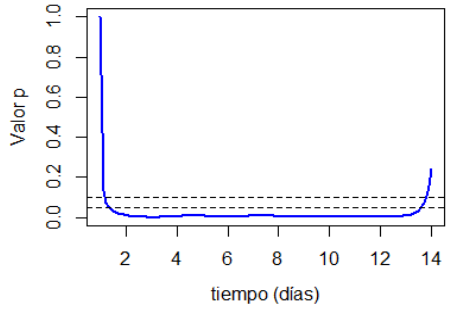
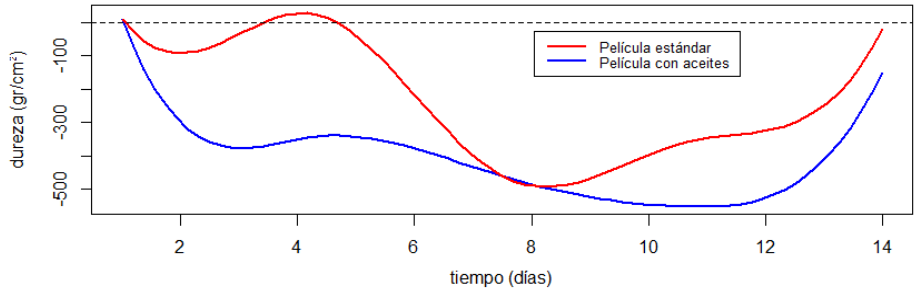



**Gráfico 5.50:** Funciones de efectos $\widehat{\alpha_1}$ y $\widehat{\alpha_2}$ sobre la dureza de papayas sin películas (control), al aplicarse películas comestibles estándar y películas comestibles con aceites esenciales respectivamente. Es deseable que el efecto sea positivo.



Observando los gráficos 5.48 y 5.49 correspondientes a la función estadístico de prueba $F(t)$ y su respectiva función $P_{VALUE}$ se observa que durante todo el experimento hasta el día trece se encuentran diferencias significativas en la dureza del fruto al aplicársele películas comestibles.

El efecto de la aplicación de películas comestibles estándar sobre la dureza del fruto es casi siempre superior que si se aplicase películas comestibles con aceites esenciales (ver gráfico 5.50) siendo esto indicio de que la inclusión de aceites esenciales en la película no tiene el efecto esperado.

La hipótesis nula de la prueba de hipótesis funcional que compara los dos tipos de película comestibles, jamás es rechazada por lo que se puede concluir que la aplicación de aceites esenciales no tiene efectos favorables en la dureza del fruto.

**Gráfico 5.51:** Función $T(t)$ análoga al estadístico de prueba T para diferencia de medias. Las varianzas se supusieron desiguales.

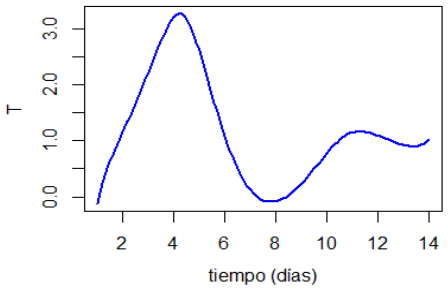
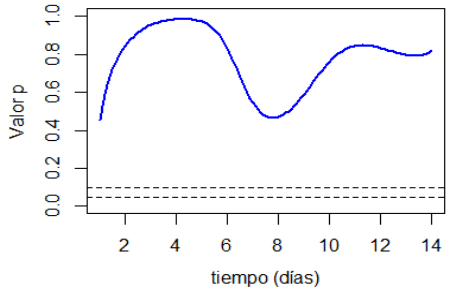



**Gráfico 5.52:** Función $P_{VALUE}$ análoga al valor p o nivel de significancia de la prueba T para todo $t$.



- ***Color***

Durante los primeros 6 días del experimentos la Suma Cuadrática de Regresión fue cero por lo que la Potencia de Explicación del Modelo fue nula para este período tiempo donde luego esta aumentó tomando entre los días 9 y 10 su valor máximo de 68.24%.

**Gráfico 5.53:** Funciones correspondientes a las Sumas Cuadráticas de Regresión, Error y Total

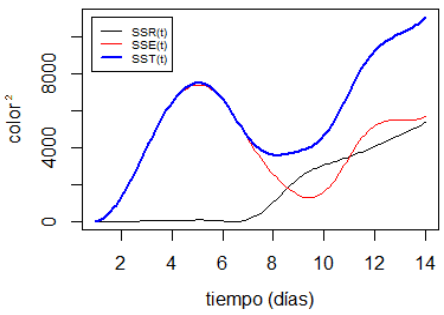



**Gráfico 5.54:** Función $\mathrm{RSQ}(t)$ análoga al Coeficiente de Determinación del Modelo $R^2$.

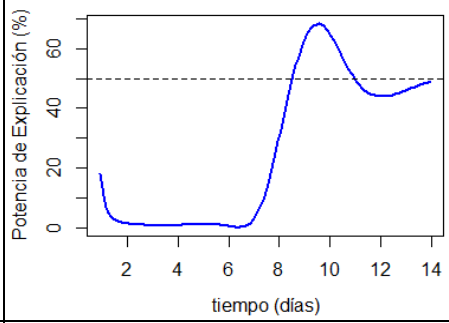



**Gráfico 5.55:** Función $F(t)$ análoga al estadístico de prueba F para el Análisis de Varianza.

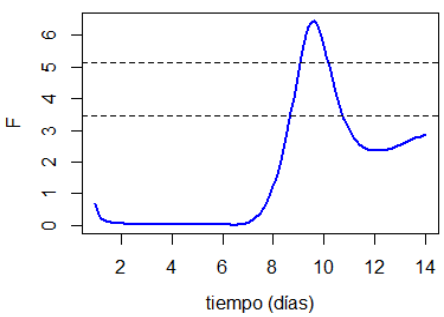



**Gráfico 5.56:** Función $P_{VALUE}$ análoga al valor p o nivel de significancia de la prueba F para todo $t$.

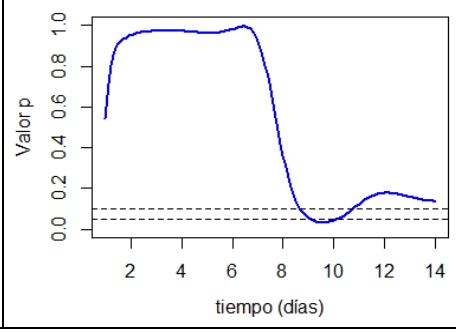
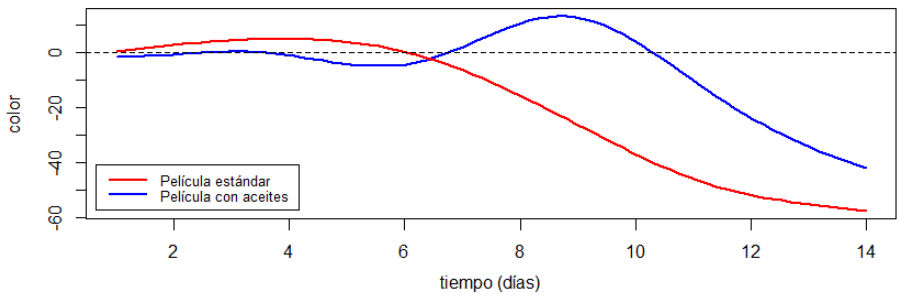



**Gráfico 5.57:** Funciones de efectos $\widehat{\alpha_1}$ y $\widehat{\alpha_2}$ sobre el color de papayas sin películas (control), al aplicarse películas comestibles estándar y películas comestibles con aceites esenciales respectivamente. Es deseable que el efecto sea negativo.



Es para este mismo período de tiempo que tanto la función $F(t)$ y $P_{VALUE}$ se encuentran en zona de rechazo por lo que únicamente en estos días (entre el noveno y décimo) es donde se encuentran diferencias significativas sobre el color, al aplicarse películas comestibles en los frutos. Véanse gráficos 5.55 y 5.56.

Se puede comprobar en el gráfico 5.57 que los efectos de la aplicación de ambos tipos de películas comestibles son nulos durante los primeros seis días del experimento, donde después de esto es el efecto correspondiente a películas comestibles estándar es negativo lo cual es de deseable.

Aplicando la prueba funcional de diferencias de medias (ver gráfico 5.59) encontramos que nunca se puede consideran que la aplicación de aceites esenciales en la película es favorable en el color del fruto, como indicador de la madurez del fruto.

**Gráfico 5.58:** Función $T(t)$ análoga al estadístico de prueba T para diferencia de medias. Las varianzas se supusieron desiguales.

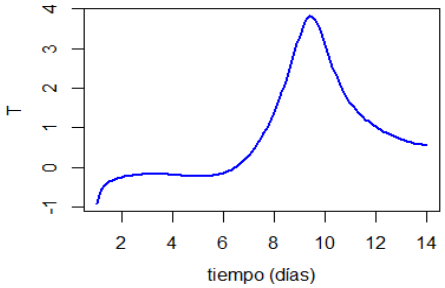
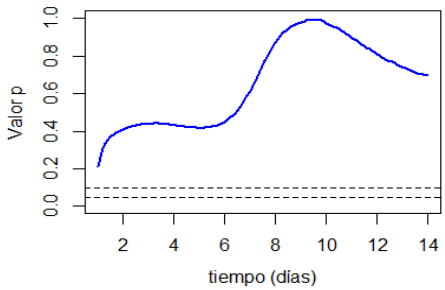



**Gráfico 5.59:** Función $P_{VALUE}$ análoga al valor p o nivel de significancia de la prueba T para todo $t$.



# CONCLUSIONES

- Se pudo comprobar la efectividad de las películas comestibles al retardar el proceso de maduración en las papayas, retrasando su pérdida de peso, previniendo el aumento del nivel pH, reduciendo los sólidos solubles y conservando la tonalidad oscura en el fruto. De quererse conservar la dureza de la papaya sería recomendable no aplicar ningún tipo de película comestible.

- La aplicación de películas comestibles reduce significativamente la variabilidad en la pérdida de peso y en la dureza del fruto, por lo que si fuese de interés predecir la pérdida promedio de peso o dureza del fruto en un instante de tiempo $t$, de no aplicarse algún tipo de revestimiento se tendría mayor incertidumbre.

- Para papayas sin algún tipo de revestimiento, la pérdida de peso se acelera durante los primeros días del experimento, mientras que para las restantes características (pH, sólidos solubles, dureza y color) la mayor parte de la pérdida total se realiza a partir de la mitad del experimento. En general, la aplicación de películas comestibles

produce que las características aumentan o disminuyan de manera más suave, es decir que las funciones tengan menos aspereza.

- La película comestible estándar resultó siempre mejor que la película comestible experimental (con aceites esenciales) al momento de reducir los cambios en los indicadores de interés que son evidencia de la maduración del fruto.

- En la prueba funcional para diferencia de medias se encontró significativamente mejor a la película experimental, al conservar el peso únicamente durante el primer día del experimento. Esto no tiene mayor relevancia, dado que para este período el fruto aún no se encuentra apto para el consumo.

- No es de desacreditar la efectividad de las películas comestibles con aceites esenciales, pues si el lugar de almacenamiento del fruto es ciertamente húmedo, esto aumentará la probabilidad de que aparezcan pudriciones fungosas en las papayas y he aquí es donde la efectividad de esta película es válida, además que siempre será mejor en la conservación de la madurez del fruto comparado sino se aplicase ningún tipo de película.

# RECOMENDACIONES

- Disponer de los recursos necesarios (laboratorios, instrumentos, equipos, personal, etc.) para realizar el experimento anterior con un mayor número de réplicas para así obtener resultados con una mayor afijación.

- Probar diferentes condiciones experimentales del entorno donde se realiza el experimento, tales como ambientes cerrados no refrigerados y/o ambientes al aire libre, esperándose diferentes resultados para la efectividad de las películas bajo estas otras condiciones ambientales.

- Dado que las mediciones no fueron realizadas a las mismas unidades experimentales durante el experimento porque los frutos eran cortados, se sugiere utilizar otros métodos para relacionar las unidades entres tiempo y tiempo; como sugerencia se podría considerar la parcela de la cual el fruto ha sido cosechado entre otros.

- Con el objetivo de medir el color es recomendable utilizar una cámara y fotografiar toda un área del fruto en lugar de realizar una o varias

mediciones puntuales con un colorímetro pues los frutos tienen diferentes pigmentaciones por área.

La fotografía del área puede ser sometida a un Tratamiento Digital de Imágenes, con el cual se puede calcular el color promedio del área así como eliminar colores aberrantes, por ejemplo manchas. Estas, y más herramientas ya se encuentran debidamente implementadas en diferentes software matemáticos y estadísticos como Matlab y R.

- Experimentar con otros agentes naturales en películas comestibles que eviten la contaminación fungosa del fruto y que sea igual o más efectiva en la conservación del fruto que una película comestible estándar.

# ANEXOS

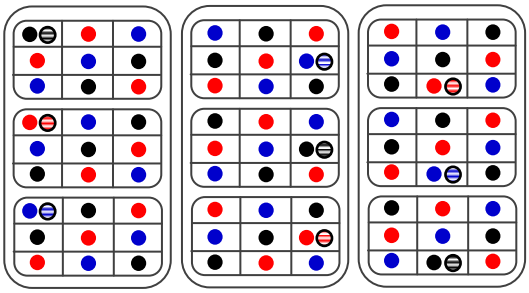
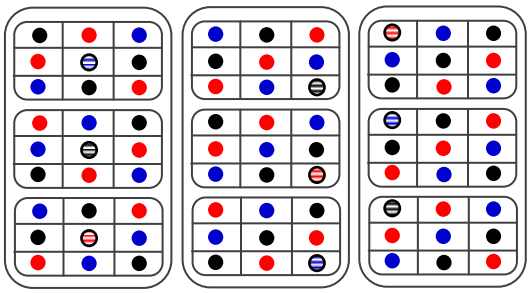
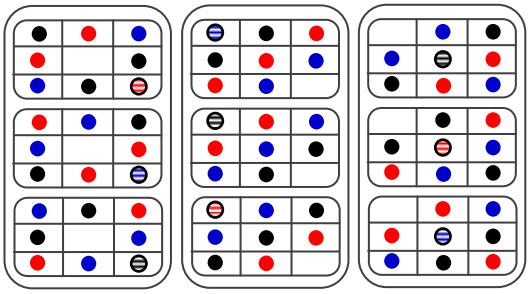
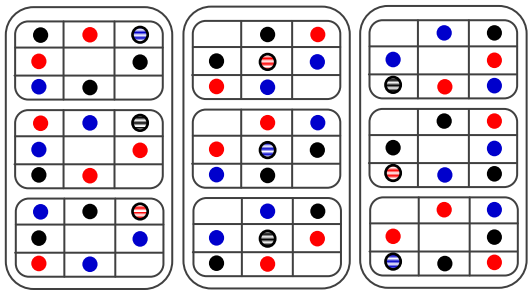
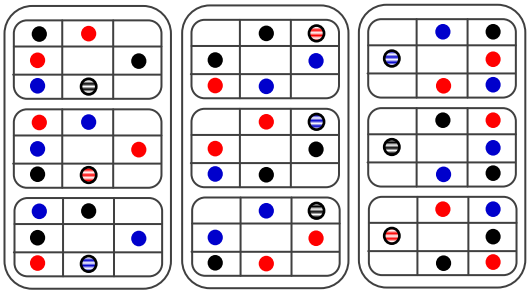
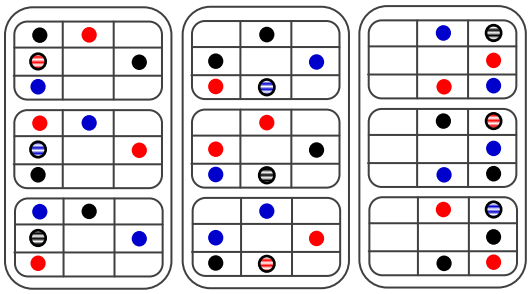
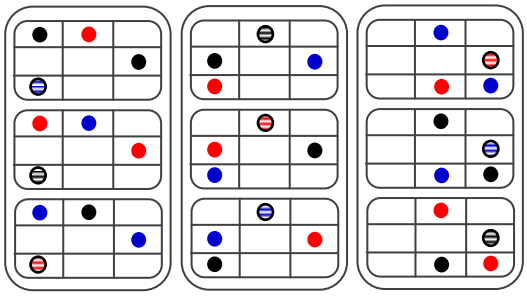
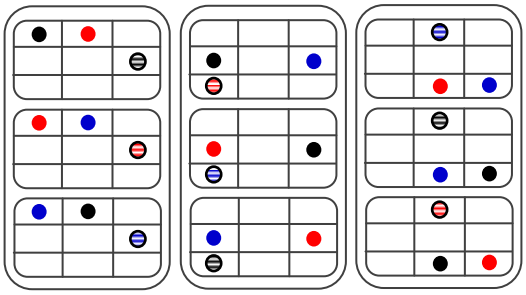
# Anexo I

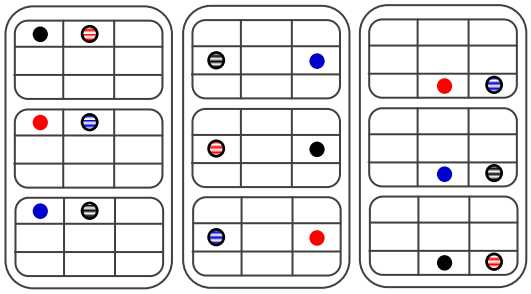
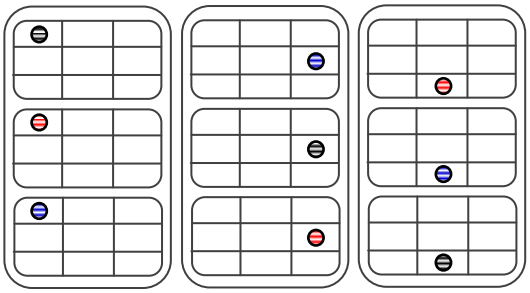
Pesos en gramos de doce papayas medidas en once ocasiones durante los quince días del experimento piloto.

| | Día 1 | Día 2 | Día 3 | Día 4 | Día 7 | Día 8 | Día 9 | Día 10 | Día 11 | Día 14 | Día 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Papaya 1** | 390,83 | 390,62 | 390,60 | 389,59 | 387,96 | 387,51 | 387,12 | 386,80 | 386,24 | - | - |
| **Papaya 2** | 386,21 | 386,10 | 386,30 | 384,54 | 383,18 | 382,63 | 382,22 | 381,76 | 381,26 | 380,62 | 379,56 |
| **Papaya 3** | 438,18 | 437,99 | 437,98 | 436,52 | 434,90 | - | - | - | - | - | - |
| **Papaya 4** | 382,08 | 381,80 | 381,86 | 380,31 | 378,87 | 378,24 | - | - | - | - | - |
| **Papaya 5** | 432,14 | 431,82 | 431,89 | 430,30 | 428,85 | 428,37 | 427,93 | 427,63 | 427,08 | 425,68 | 425,30 |
| **Papaya 6** | 392,40 | 392,21 | - | - | - | - | - | - | - | - | - |
| **Papaya 7** | 345,81 | 345,58 | 345,64 | 344,82 | 344,06 | 343,80 | 343,70 | - | - | - | - |
| **Papaya 8** | 444,17 | 443,98 | 444,12 | 443,35 | 442,11 | 441,96 | 441,63 | 441,40 | 441,00 | 439,80 | - |
| **Papaya 9** | 382,28 | 382,05 | 382,02 | 380,92 | 379,67 | 379,20 | 378,97 | 378,70 | - | - | - |
| **Papaya 10** | 465,39 | 465,15 | 465,18 | - | - | - | - | - | - | - | - |
| **Papaya 11** | 475,31 | 475,12 | 475,20 | 474,66 | - | - | - | - | - | - | - |
| **Papaya 12** | 447,09 | 446,79 | 446,87 | 446,31 | 445,07 | 444,75 | 444,53 | 444,37 | 444,06 | 443,05 | 442,63 |

# Anexo II

Representación gráfica de la selección aleatoria de papayas día a día durante el experimento. Obsérvese las papayas sin película (negro), con película (rojo) y con película con aceites esenciales (azul). Las papayas seleccionadas para ser analizadas, se representar con contorno negro y tramado.

| Día 1 | Día 2 |
|---|---|
|  |  |
| **Día 3** | **Día 4** |
|  |  |
| **Día 5** | **Día 6** |
|  |  |
| **Día 7** | **Día 8** |
|  |  |

| Día 9 | Día 10 |
| --- | --- |
|  |  |

# BIBLIOGRAFÍA

[1] ***Abramovich, F., Antoniadis, A., Sapatinas, T. y Vidakovic, B.*** (2002). Optimal testing in functional analysis of variance models. Georgia Institute of Technology ISYE Statistics Technical Report.

[2] ***Almeida A., Reis J. D., Santos D., Vieira, T. y da Costa, M.*** (2011), Study of preservation of papaya (Carica papaya L.) associated with the application of edible films, Universidade Federal de Sergipe (UFS), São Cristóvão, Sergipe (SE), Brasil.

[3] ***de Boor, C.*** (2001), A Practical Guide to Splines, Springer series in statistics, New York - USA.

[4] ***Hoerl, Arthur E.*** (1962), Application of ridge analysis to regression problems, Chemical Engineering Progress, 58, 54-59.

[5] ***Hooker, G.*** (2010), Introduction to Functional Data Analysis, International Workshop on Statistical Modeling, Cornell University, Ithaca, NY - USA.

[6] ***http://www.functionaldata.org***, actualizado a Septiembre de 2009 y consultado a Septiembre de 2011.

[7] ***http://solidossolublesdenana.blogspot.com***, actualizado a Marzo de 2012 y consultado a Mayo de 2012.

[8] ***http://www.statistics.com/index.php?page=glossary&term_id=766,*** actualizado a Agosto de 2011 y consultado a Septiembre de 2011.

[9] ***Kutner M., Nachtsheim C. y Neter J.,*** *Applied Linear Regression Models*, 4th edition, McGraw-Hill Irwin, 2004.

[10] ***Press, W. H., Teukolsky, S. A., Vetterling, W. T. y Flannery, B. P.*** (1999), Numerical recipes in C, Segunda edición, Cambridge, Cambridge University Press.

[11] ***Ramsay, J.O. y Silverman, B.W.*** (2005), Functional data analysis, Segunda edición, Springer series in statistics, New York - USA.

[12] ***SAS/STAT(R)*** (2012), Manual de Usuario, Versión 9.2, Segunda edición. Disponible en la web en http://support.sas.com/documentation/cdl/en /statug/63033/HTML/default/viewer.htm#statug_introcom_a0000000525.htm

[13] ***Satterwaite, F. E.*** 1946. An approximate distribution of estimates of variance components. Biometrics Bull 2:110–4.

[14] ***Stein, Charles M.*** (1956), “Inadmissibility of the usual estimator for the mean of a multivariate distribution”, *Proc. Third Berkeley Symp*, **1**:197–206.

[15] ***Tychonoff, Andrey N.*** (1943), "Об устойчивости обратных задач [On the stability of inverse problems]". *Doklady Akademii Nauk SSSR,* **39** (5): 195–198.

[16] ***Zurita, G.*** (2010). “Probabilidad y Estadística, Fundamentos y Aplicaciones”, Segunda Edición, Instituto de Ciencias Matemáticas ESPOL, Guayaquil, Ecuador.