



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

**FACULTAD DE CIENCIAS**

**COMPARACIÓN A TRAVÉS DE EJEMPLOS ENTRE  
REGRESIÓN LINEAL Y REGRESIÓN BAYESIANA**

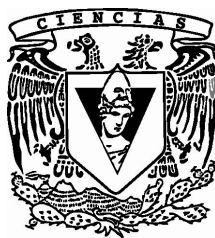
**T E S I S**

**QUE PARA OBTENER EL TÍTULO DE:**

**M A T E M Á T I C O**

**P R E S E N T A:**

**JORGE ANTONIO LÓPEZ CROTTE**



**DIRECTOR DE TESIS:  
M. en C. MARÍA DEL PILAR ALONSO REYES  
2010**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Datos del jurado de tesis.

1	Propietario	M. en C.	<b>José Antonio</b>	<b>Flores</b>	<b>Díaz</b>
		Grado	Nombre(s)	Apellido Paterno	Apellido Materno
2	Propietario	Act.	<b>Francisco</b>	<b>Sánchez</b>	<b>Villareal</b>
		Grado	Nombre(s)	Apellido Paterno	Apellido Materno
3	Propietario	M. en C.	<b>María del Pilar</b>	<b>Alonso</b>	<b>Reyes</b>
	Tutora	Grado	Nombre(s)	Apellido Paterno	Apellido Materno
4	Suplente	M. en C.	<b>Hugo</b>	<b>Villaseñor</b>	<b>Hernández</b>
		Grado	Nombre(s)	Apellido Paterno	Apellido Materno
5	Suplente	M.I.O.	<b>María Isabel</b>	<b>Escalante</b>	<b>Membrillo</b>
		Grado	Nombre(s)	Apellido Paterno	Apellido Materno

## DEDICATORIAS

*Esta tesis está dedicada a mi madre, a quien agradezco de todo corazón su amor y comprensión, pues a lo largo de mi vida me ha alentado a seguir adelante y siempre me ayudó en todo lo que pudo; le agradezco por haber entendido mis ausencias y mis malos ratos, así como compartir mis momentos de felicidad y de tristeza.*

*En todo momento la llevo conmigo.*

## AGRADECIMIENTOS

*Este trabajo es un esfuerzo en el cual, directa o indirectamente, participaron varias personas leyendo, opinando, corrigiendo, teniéndome paciencia, dando ánimo, acompañando en los momentos de crisis y en los momentos de felicidad.*

*Agradezco a Dios por llenar mi vida de dichas y bendiciones.*

*Agradezco a mi hermano por la compañía y el apoyo que me brinda. Se que cuento con el siempre.*

*A mi padre, que a pesar de la distancia siempre ha estado pendiente mí.*

*Agradezco a mis amigos por su confianza y lealtad.*

*Agradezco a mi país porque espera lo mejor de mí.*

*Agradezco a mis maestros por su disposición y ayuda brindadas.*

*Agradezco especialmente a mi directora de tesis por haber confiado en mi persona, por la paciencia y por el seguimiento de este trabajo.*

*Gracias a todos.*

# Índice

Índice .....	1
Introducción al análisis de regresión simple.....	3
Comparación a través de ejemplos entre regresión lineal y regresión bayesiana.....	4
1ra Parte: Regresión clásica.....	4
1.1 Introducción a la regresión clásica .....	4
1.2 Regresión lineal simple .....	6
1.3 Formulación del modelo.....	6
1.4 Estimación de los parámetros del modelo .....	7
1.4.1 Método de máxima verosimilitud.....	8
1.4.2 Método de mínimos cuadrados.....	9
1.4.2.1 Propiedades de los estimadores por mínimos cuadrados y el modelo de regresión ajustado .....	11
1.4.3 Estimación de $\sigma^2$ .....	13
1.5 Distribución de los estimadores mínimos cuadrados .....	14
1.5.1 Intervalo de confianza para $\beta_0$ , $\beta_1$ y $\sigma^2$ .....	14
1.5.2 Intervalo de la respuesta media .....	16
1.5.3 Prueba de hipótesis para la pendiente $\beta_1$ y la intersección $\beta_0$ .....	18
1.6 Coeficiente de determinación y coeficiente de correlación.....	21
1.6.1 Interpretación del coeficiente de correlación.....	23
1.6.2 El $R^2$ ajustado .....	24
1.7 Predicción en regresión lineal simple.....	25
1.7.1 Predicción de nuevas observaciones.....	25
2da Parte: Regresión bayesiana .....	27
2.1 Teoría de la decisión.....	27
2.2 Introducción a la estadística bayesiana.....	28
2.3 Inferencia bayesiana: distribuciones a priori y posteriori.....	30
2.4 Estimación de Bayes.....	33
2.5 Familias conjugadas .....	34
2.6 Distribuciones a priori no informativas .....	35
2.7 Regla de Jeffreys .....	36
2.8 Modelo bayesiano de regresión lineal simple.....	36
2.8.1 Regresión por mínimos cuadrados .....	36

2.8.2	Forma alternativa de la recta de mínimos cuadrados .....	37
2.8.3	Estimación de la varianza por mínimos cuadrados .....	38
2.8.4	Supuestos de la recta de regresión.....	38
2.9	Teorema de Bayes para el modelo de regresión.....	39
2.9.1	La función de verosimilitud conjunta para $\alpha_x$ y $\beta_1$ .....	40
2.9.2	La a priori conjunta para $\alpha_x$ y $\beta_1$ .....	42
2.9.3	Eligiendo las funciones a priori normales de $\alpha_x$ y $\beta_1$ .....	43
2.9.4	La función posteriori conjunta para $\beta_0$ y $\beta_1$ .....	43
2.9.5	Intervalo de credibilidad bayesiano para la pendiente .....	45
2.9.6	Intervalo de confianza frecuentista para la pendiente .....	46
2.9.7	Pruebas de hipótesis bayesianas .....	46
2.9.8	Prueba de hipótesis de una cola para la pendiente.....	46
2.9.9	Prueba de hipótesis de dos colas para la pendiente .....	47
2.9.10	Similitudes entre pruebas de hipótesis frecuentista y bayesiana .....	48
2.9.11	Diferencias entre pruebas de hipótesis frecuentista y bayesiana.....	48
2.9.12	El principio de máxima verosimilitud .....	48
2.10	Distribución de la predicción para futuras observaciones.....	49
2.10.1	Distribución de la predicción.....	50
2.10.2	Intervalo de credibilidad para la predicción .....	52
3ra	Parte: Ejemplos de regresión lineal múltiple con enfoque clásico .....	53
3.1	Ejemplo 1.....	53
3.2	Ejemplo 2.....	73
3.3	Ejemplo 3.....	83
4ta	Parte: Ejemplos de regresión lineal simple con enfoque bayesiano .....	92
4.1	Ejemplo 1.....	94
4.2	Ejemplo 2.....	101
4.3	Ejemplo 3.....	107
	Resultados y conclusiones .....	114
	Conclusiones de los resultados .....	115
	Conclusiones generales.....	117
	Bibliografía.....	119



## **Introducción al análisis de regresión simple**

En la naturaleza existen una variedad enorme de fenómenos que se pueden estudiar a través de modelos matemáticos, para esto se pueden utilizar los de regresión, pues ellos describen y/o representan la relación entre dos variables que están asociadas. Es una de las técnicas estadísticas más desarrollada y más utilizada. Modelar implica el desarrollo y la aplicación de una ecuación matemática que describe, de alguna manera, el comportamiento de una variable aleatoria, en este caso, se referirá a la ecuación de una recta con una ordenada al origen y una pendiente, aunque se pudieran hacer diversos ejemplos con diferentes tipos de curvas polinomiales. En estadística se proponen representaciones matemáticas y se supone que se ajustan o se adecuan a los datos. Se estima la media y la constante de cambio.

Por lo tanto el objetivo del análisis de regresión es estimar el valor promedio de una variable, llamada dependiente, con base en los valores de una o más variables adicionales, denominadas explicativas. Esta técnica ha cobrado popularidad debido al gran número de paquetes estadísticos que lo incluyen y por ser un proceso que se adapta a un sinnúmero de aplicaciones científicas y ejecutivas que permite la toma de decisiones. En esta tesis, el mejor ajuste de los modelos estará determinado por el análisis de regresión lineal simple, utilizando dos métodos estadísticos.

La interpretación de la recta, así como los métodos que se utilizan para estimar los parámetros, dependerán del tipo de variables y de la distribución de las mismas.

# **Comparación a través de ejemplos entre regresión lineal y regresión bayesiana**

## **1ra Parte: Regresión clásica**

### **1.1 Introducción a la regresión clásica**

Uno de los objetivos de la ciencia es llegar a conocer y comprender los sucesos naturales y la causa de los mismos. Tanto los fenómenos de la naturaleza como los humanos son complejos: la ocurrencia o no de los mismos depende de múltiples factores. Tratar de explicar la existencia y/o intensidad de determinados eventos implica conocer con qué otros está asociado. En su caso, en un estado posterior se podría analizar las causas que lo originan. En ocasiones se requiere conocer en qué circunstancias se producen algunos hechos, en otras palabras, se pretende saber por qué alcanzan determinada magnitud y cuáles son los cambios de los mismos.

Alrededor de 1860, Sir Francis Galton<sup>1</sup>, estudiando la relación coexistente entre las alturas de padres e hijos, llegó al primer modelo conocido ahora como de "regresión". De hecho, este nombre se debe al propio Galton, quien describía esta relación entre dichas alturas como una "regresión a la mediocridad" o llamado por otros autores como la "ley de la regresión filial", indicando con esto que los padres altos tenían hijos ya no tan altos y los padres bajos tenían hijos ya no tan bajos, existiendo una tendencia a regresar hacia los valores medios.

La regresión es una técnica estadística que se utiliza para investigar y modelar las relaciones entre variables, con el propósito de usar la información que proporciona una de ellas para tratar de conocer en forma aproximada el comportamiento de la otra. El beneficio que se deriva de llevar a cabo un procedimiento como el expuesto anteriormente es de diversos órdenes, por ejemplo, puede ser más económico o práctico observar (medir) una característica que otra, en tal sentido sería conveniente poder "predecir" valores de la variable que presenta problemas con base en la observación de la otra. Por ejemplo, en un estudio se desea conocer el grado de conocimiento político

---

<sup>1</sup> Matemático inglés nacido en Sparkbrook, Inglaterra, el 16 de Febrero de 1822 y fallecido en Grayshott House, Haslemere, Surrey, Inglaterra, el 17 de Enero de 1911.

de las personas, el cual puede ser difícil de medir, sin embargo, otra variable muy relacionada puede medirse tal como, años de educación, edad, o ingreso.

En otras palabras, se puede decir que el fin primordial de la regresión es el de tratar de explicar los cambios que se producen en una característica que se especifica como dependiente con base a las diferencias que se producen en otra u otras variables. Pero, se debe ser consciente de que al hablar de regresión, explicación o ajuste, no se está, necesariamente, hablando de causalidad. Esto último es de suma importancia, ya que dos variables pueden con un alto grado, estar relacionadas, pero no ser una causal de la otra o viceversa.

La explicación de determinados comportamientos o características puede posibilitar el hacer predicciones acerca de la ocurrencia o no del hecho que se está estudiando.

La regresión múltiple permite analizar tanto las contribuciones individuales como las colectivas de un conjunto de variables independientes a la explicación de los cambios que se producen en la dependiente, es decir, permite analizar cómo afectan el primer grupo, de forma conjunta, el comportamiento de la segunda. También deja estudiar, de forma pormenorizada, la influencia de cada una de las variables explicativas en la de objeto de estudio.

Los propósitos que se plantean o se conceptualizan en la regresión pueden ser muy ambiciosos, por lo que, una mala utilización de la misma también puede conducir a conclusiones erróneas que no se ajustan a la realidad de los datos o de la población a la que se trata de representar.

En ocasiones el uso del modelo de regresión es útil por la estimación de los parámetros en algunas ciencias son de interés particular, como por ejemplo, el coeficiente de elasticidad en la economía.

También se puede utilizar en otras ramas de la ciencia, como por ejemplo, en medicina se puede realizar un estudio para cuantificar la relación entre la pérdida diaria de lípidos con la pérdida de energía en las heces de niños con fibrosis quística. Otro ejemplo sería en las ciencias marinas, en donde se puede conocer la forma de la relación que existe

entre la longitud de una especie y su edad. También, en vulcanología se pudiera predecir el intervalo de la siguiente erupción de un volcán a partir de la duración de la última erupción, además de otras múltiples aplicaciones.

## 1.2 Regresión lineal simple

El hecho de que exista una relación entre dos variables significa que el conocimiento que se tenga del valor de una de ellas aporta algún tipo de información acerca del resultado obtenido en la segunda variable.

En un análisis de regresión lineal simple existe una variable respuesta o dependiente  $y$  y una explicativa o independiente  $x$ . La finalidad es obtener una función sencilla de la segunda, que sea capaz de ajustar la variación de la primera lo más posible. Como los valores observados de la variable dependiente difieren generalmente de los que predicen la función, la recta más eficaz será aquella que describa a la de respuesta con la menor diferencia entre los valores observados y predichos (lo que es llamado el error).

## 1.3 Formulación del modelo

La ecuación de regresión lineal simple es una función lineal:

$$y = \beta_0 + \beta_1 x + \varepsilon,$$

donde  $\beta_0$  es el valor de la ordenada, es decir, es donde la recta se intercepta con el eje vertical,  $\beta_1$  es el coeficiente de regresión poblacional (pendiente de la línea recta) y  $\varepsilon$  es el error.

En la teoría, se supone el hecho de que los errores  $\varepsilon$  tienen media cero y varianza desconocida pero constante  $\sigma^2$ , adicionalmente se asumirá que estos no están correlacionados, lo cual significa que el valor de cualquiera de ellos no depende del resto.

Es conveniente ver a la variable  $x$  como la que es controlada por el analista de datos y que se mide con poco error, mientras que la variable  $y$  es aleatoria, esto es, que tiene una distribución de probabilidad para  $y$  en cada posible valor de  $x$ . La media de esta distribución es:

$$E(y|x) = \beta_0 + \beta_1 x$$

y la varianza es:

$$V(y|x) = V(\beta_0 + \beta_1 x + \varepsilon) = \sigma^2.$$

Así pues, la media de la variable dependiente es una función lineal de la variable independiente, aunque la varianza de  $y$  no depende del valor de  $x$ . Además, dado que los errores no tienen correlación, las respuestas mantienen esta propiedad.

La intención es encontrar los factores numéricos  $\beta_0$  y  $\beta_1$ , también nombrados coeficientes de regresión, que producen la función lineal que mejor se ajusta a los datos (esta será denotada como  $\hat{y}$ ). Entonces, se emplean los pares ordenados de puntos para encontrar los coeficientes, de tal manera que la línea de regresión aproxime los valores observados con los calculados.

Para definir los valores que mejor se ajusten a los datos se puede ocupar el error de predicción, es decir, la diferencia de  $y$  con respecto a  $\hat{y}$ . Este se llama también residuos o residuales (cuando  $y$  no es predicho por  $\hat{y}$ ).

## 1.4 Estimación de los parámetros del modelo

Existen tres parámetros que se deben estimar en la ecuación de regresión lineal simple, y estos son los coeficientes de la recta de regresión,  $\beta_0$  y  $\beta_1$ , además de la varianza  $\sigma^2$  de los errores, que como uno de los supuestos, están distribuidos normalmente.

El cálculo de estimadores para estos parámetros puede hacerse por diferentes métodos, siendo los más utilizados el de máxima verosimilitud y el de mínimos cuadrados.

### 1.4.1 Método de máxima verosimilitud

Los estimadores máximos verosímiles surgen de maximizar la función de verosimilitud

$$f(\beta_0, \beta_1, \sigma^2; Y = (y_1, \dots, y_n)) = \prod_{i=1}^n f(y_i | x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (y_i - \beta_0 - \beta_1 x_i)^2\right),$$

donde  $n$  es el número total de datos.

A través de las derivadas parciales de esta función se obtienen las siguientes ecuaciones

$$\sum_{i=1}^n y_i - \beta_0 - \beta_1 x_i = 0,$$

$$\sum_{i=1}^n y_i x_i - \beta_0 x_i - \beta_1 x_i^2 = 0,$$

$$\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 - n\sigma^2 = 0.$$

Estas tres condiciones son llamadas las ecuaciones normales del método de máxima verosimilitud.

Y las soluciones de éstas son:

$$\hat{\beta}_{0,MV} = \bar{y} - \hat{\beta}_{1,MV} \bar{x},$$

$$\hat{\beta}_{1,MV} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

$$\sigma_{MV}^2 = \frac{\sum_{i=1}^n (y_i - \hat{\beta}_{0,MV} - \hat{\beta}_{1,MV}x_i)^2}{n},$$

donde

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

### 1.4.2 Método de mínimos cuadrados

Muchas rectas pueden satisfacer la función del valor medio de  $y$ , dependiendo de los valores de  $\beta_0$  y  $\beta_1$ . La mejor línea será la que pase por el centro de los datos con el mejor ajuste. Esta es aquella cuya diferencia entre cada valor observado  $y_i$  y la recta  $\hat{y}_i = \beta_0 + \beta_1 x_i$  sea la más pequeña, es decir, cuando  $|y_i - \hat{y}_i|$  sea la menor de todas las distancias. Entonces el ajuste conveniente es aquel con la diferencia total menor y se obtiene minimizando la suma de los errores al cuadrado, es decir:

$$S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \beta_0 + \beta_1 x_i)^2,$$

tiene que ser mínimo, para lo cual se deriva esta ecuación en términos de  $\beta_0$  y  $\beta_1$ , así se pueden encontrar los estimadores de los mismos. Teniendo en cuenta que  $n$  es el número de los datos.

Las soluciones del método son:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Entonces se reescribe la función de regresión lineal simple como:

$$\hat{y}_i = \beta_0 + \beta_1 x_i.$$

La ecuación anterior da un punto estimado de la media de  $y_i$  para un dato en particular  $x_i$ .

Las expresiones

$$s_{xx} = \sum_{i=1}^n x_i^2 - n\bar{x}^2 = \sum_{i=1}^n (x_i - \bar{x})^2,$$

$$s_{xy} = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

son llamadas, la suma de cuadrados de  $x_i$  y la suma de productos cruzados de  $x_i$  y  $y_i$ , correspondientemente.

Por lo que la ecuación para estimar a  $\beta_1$  se puede escribir como sigue:

$$\beta_1 = \frac{s_{xy}}{s_{xx}}.$$

Como ya se había mencionado, a la diferencia de los valores observados  $y_i$  y los correspondientes calculados  $\hat{y}_i$  se les llama residuos o residuales. Matemáticamente se denotan a los residuos como sigue:

$$e_i = y_i - \hat{y}_i = y_i - (\beta_0 + \beta_1 x_i), \quad i = 1, 2, \dots, n.$$

Estos juegan un papel muy importante dentro de la obtención del mejor modelo de regresión ajustado.



### 1.4.2.1 Propiedades de los estimadores por mínimos cuadrados y el modelo de regresión ajustado

Los estimadores obtenidos por el método de mínimos cuadrados,  $\hat{\beta}_0$  y  $\hat{\beta}_1$ , tienen varias propiedades estadísticas muy importantes. Primeramente se sabe que los coeficientes son combinaciones lineales de las observaciones  $y_i$ , es decir,

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} = \sum_{i=1}^n c_i y_i,$$

donde  $c_i = (x_i - \bar{x})/S_{xx}$  para  $i = 1, 2, \dots, n$ .

Considerando la propiedad de sesgo para los estimadores  $\hat{\beta}_0$  y  $\hat{\beta}_1$ , se tiene entonces que:

a) Para  $\hat{\beta}_1$  la esperanza es

$$E(\hat{\beta}_1) = \beta_1,$$

y su varianza está dada por

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

b) Para  $\hat{\beta}_0$  la esperanza es

$$E(\hat{\beta}_0) = \beta_0,$$

y su varianza será

$$V(\beta_0) = \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right).$$

Existen varias propiedades sobre la línea de regresión ajustada, las cuales son:

1. La suma de los residuos en cualquier modelo de regresión que contiene a la intersección  $\beta_0$  es siempre cero, es decir,

$$\sum_{i=1}^n (y_i - \hat{y}_i) = \sum_{i=1}^n e_i = 0.$$

2. La suma de los residuos al cuadrado,

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2 \text{ es mínima.}$$

3. La suma de los valores observados  $y_i$  es igual a la suma de los ajustados  $\hat{y}_i$ ,

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i.$$

4. La recta de regresión obtenida por mínimos cuadrados pasa por el centro de los datos, es decir, por el punto  $(\bar{y}, \bar{x})$ .

5. La suma de los residuos ponderados por el correspondiente valor de la variable explicativa siempre es igual a cero, esto es,

$$\sum_{i=1}^n x_i e_i = 0.$$

6. La suma de los residuos ponderados por el correspondiente valor ajustado siempre es igual a cero, es decir,

$$\sum_{i=1}^n \hat{y}_i e_i = 0.$$

### 1.4.3 Estimación de $\sigma^2$

Un estimador de  $\sigma^2$  es requerido para las pruebas de hipótesis y para construir intervalos de confianza para el modelo de regresión. Idealmente se pretendería que esta estimación no dependiera de la adecuación de los modelos ajustados. Esto sólo es posible cuando hay bastantes observaciones en  $y$  para al menos un valor de  $x$  o cuando la información relativa a  $\sigma^2$  está disponible. Cuando este enfoque no puede ser utilizado, la estimación de  $\sigma^2$  es obtenida desde los residuos o de la suma de los cuadrados de los errores.

$$\begin{aligned}SS_E &= \sum_{i=1}^n e_i^2 \\ &= \sum_{i=1}^n (y_i - \hat{y}_i)^2.\end{aligned}$$

Una fórmula conveniente para  $SS_E$  puede ser encontrada si se sustituye  $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$  en la última ecuación

$$SS_E = \sum_{i=1}^n y_i^2 - n\bar{y}^2 - \hat{\beta}_1 S_{xy},$$

pero

$$S_{yy} = \sum_{i=1}^n y_i^2 - n\bar{y}^2 = \sum_{i=1}^n (y_i - \bar{y})^2.$$

Así, la suma de los cuadrados de las observaciones, es:

$$SS_E = S_{yy} - \hat{\beta}_1 S_{xy}.$$

La suma de los cuadrados de las observaciones tiene  $n-2$  grados de libertad, porque dos grados de libertad están asociados con los estimadores  $\hat{\beta}_0$  y  $\hat{\beta}_1$  que participan en la

obtención de  $\hat{y}_i$ . Ahora el valor esperado de  $SS_E$  es  $E(SS_E) = (n-2)\sigma^2$ , entonces un estimador insesgado de  $\sigma^2$  es

$$\hat{\sigma}^2 = \frac{SS_E}{n-2} = MS_E.$$

La cantidad  $MS_E$  es llamada el cuadrado medio de los errores o el cuadrado medio de los residuos. La raíz cuadrada de  $\hat{\sigma}^2$  es conocida como el error estándar de la regresión, y éste tiene las mismas unidades de medición que la variable de respuesta  $y$ . Debido a que  $\hat{\sigma}^2$  depende de la suma de los cuadrados de los residuos, alguna violación de los supuestos sobre los errores del modelo o alguna falta de información del modelo puede tener serios problemas al utilizar a  $\hat{\sigma}^2$  como un estimador de  $\sigma^2$ .

## 1.5 Distribución de los estimadores mínimos cuadrados

Para efecto de hacer inferencia en regresión, se requiere asumir que los errores  $\varepsilon_i$ , se distribuyen en forma normal e independientemente con media 0 y varianza constante  $\sigma^2$ . En consecuencia, también las  $y_i$ 's se distribuyen normalmente con media  $\beta_0 + \beta_1 x_i$  y varianza  $\sigma^2$ .

Entonces se puede establecer que  $\hat{\beta}_0$  se distribuye como  $normal\left(\beta_0, \frac{\sigma^2}{S_{xx}}\right)$  y  $\hat{\beta}_1$  se distribuye como  $normal\left(\beta_1, \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\sigma^2\right)$ .

### 1.5.1 Intervalo de confianza para $\beta_0$ , $\beta_1$ y $\sigma^2$

Uno de los supuestos del modelo de regresión es que cada error se distribuye normal e independientemente de los demás. Por lo que, la distribución de muestreo de  $\beta_0$  y de  $\beta_1$  están dadas por

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{MS_E/S_{xx}}} \quad \text{y} \quad \frac{\hat{\beta}_0 - \beta_0}{\sqrt{MS_E(1/n + \bar{x}^2/S_{xx})}}$$

las cuales son *t de Student* con  $n - 2$  grados de libertad, respectivamente. Entonces, un intervalo de confianza al  $100(1 - \alpha)\%$  para la pendiente  $\beta_1$  está dado por

$$\hat{\beta}_1 - t_{n-2}^{1-\alpha/2} \sqrt{\frac{MS_E}{S_{xx}}} \leq \beta_1 \leq \hat{\beta}_1 + t_{n-2}^{1-\alpha/2} \sqrt{\frac{MS_E}{S_{xx}}}.$$

De igual manera, un intervalo de confianza al  $100(1 - \alpha)\%$  para la intersección  $\beta_0$  es

$$\hat{\beta}_0 - t_{n-2}^{1-\alpha/2} \sqrt{MS_E \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)} \leq \beta_0 \leq \hat{\beta}_0 + t_{n-2}^{1-\alpha/2} \sqrt{MS_E \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}.$$

Estos intervalos a lo que se refieren es que si se pudieran tomar repetidas muestras del mismo tamaño y construir, por ejemplo, un intervalo al 95% de confianza para la pendiente para cada una de las muestras, entonces 95% de estos intervalos contendrán el verdadero valor de  $\beta_1$ .

La cantidad

$$se(\hat{\beta}_1) = \sqrt{\frac{MS_E}{S_{xx}}}$$

es llamada el error estándar de la pendiente  $\hat{\beta}_1$ .

Similarmente

$$se(\hat{\beta}_0) = \sqrt{MS_E \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}$$

es el error estándar de la intersección  $\hat{\beta}_0$ .

Los paquetes de computación que obtienen regresiones, usualmente les llaman a estos los errores estándares de los coeficientes de regresión.

Ahora, debido a que cada error se distribuye normal e independiente de todos los demás, la distribución muestral de

$$(n-2)MS_E / \sigma^2$$

es una *ji-cuadrada* con  $(n-2)$  grados de libertad. Así

$$P\left\{\chi_{\alpha/2, n-2}^2 \leq \frac{(n-2)MS_E}{\sigma^2} \leq \chi_{1-\alpha/2, n-2}^2\right\} = 1 - \alpha$$

y consecuentemente un intervalo de confianza al  $100(1-\alpha)\%$  para  $\sigma^2$  es

$$\frac{(n-2)MS_E}{\chi_{1-\alpha/2, n-2}^2} \leq \sigma^2 \leq \frac{(n-2)MS_E}{\chi_{\alpha/2, n-2}^2}.$$

### 1.5.2 Intervalo de la respuesta media

Un mayor uso para el modelo de regresión es poder estimar la respuesta media  $E(y)$  para un valor de  $x$  en particular. Sea  $x_0$  un punto particular de la variable explicativa para el que se quiere estimar la respuesta media, es decir,  $E(y | x_0)$ . Se asumirá que  $x_0$  toma cualquiera de los valores de la variable independiente, dentro del rango del dato original sobre  $x$  utilizado para ajustar el modelo. Un estimador insesgado de  $E(y | x_0)$  se encuentra del modelo ajustado

$$E(y | x_0) = \hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0.$$

Para obtener un intervalo de confianza al  $100(1-\alpha)\%$  de  $E(y|x_0)$ , primero hay que notar que  $\hat{y}_0$  es una variable aleatoria que está normalmente distribuida porque es una combinación lineal de las observaciones  $y_i$ . La varianza de  $\hat{y}_0$  es

$$\begin{aligned} V(\hat{y}_0) &= V(\hat{\beta}_0 + \hat{\beta}_1 x_0) \\ &= V[\bar{y} + \hat{\beta}_1(x_0 - \bar{x})] \\ &= \frac{\sigma^2}{n} + \frac{\sigma^2(x_0 - \bar{x})^2}{S_{xx}} \\ &= \sigma^2 \left[ \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right], \end{aligned}$$

pues  $Cov(\bar{y}, \hat{\beta}_1) = 0$ .

Así entonces la distribución de muestreo de

$$\frac{\hat{y}_0 - E(y|x_0)}{\sqrt{MS_E \left( \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}}$$

es una *t de Student* con  $n-2$  grados de libertad. En consecuencia, un intervalo de confianza para la respuesta media al  $100(1-\alpha)\%$  en el punto  $x = x_0$  es

$$\hat{y}_0 - t_{n-2}^{1-\alpha/2} \sqrt{MS_E \left( \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \leq E(y|x_0) \leq \hat{y}_0 + t_{n-2}^{1-\alpha/2} \sqrt{MS_E \left( \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}.$$

Nótese que el tamaño de este intervalo es una función de  $x_0$ . La longitud de éste es mínima para  $x_0 = \bar{x}$  y aumenta cuando  $|x_0 - \bar{x}| \neq 0$ . Intuitivamente esto es razonable, como sería esperar que las mejores estimaciones de  $y$  se hicieran en valores de  $x$

cercanos al centro de los datos y la precisión de la estimación tiende a deteriorarse a medida que se está más cerca del límite del espacio de las  $x$ 's.

### 1.5.3 Prueba de hipótesis para la pendiente $\beta_1$ y la intersección $\beta_0$

Además de la construcción de intervalos de confianza, a menudo se está interesado en pruebas de hipótesis para los parámetros del modelo. Es muy importante resaltar nuevamente la necesidad de que se requiere el supuesto de que los errores  $\varepsilon_i$  están distribuidos normalmente. Así, se tienen todas las condiciones para los errores, que son normalmente e independientemente distribuidos con media 0 y varianza  $\sigma^2$ , abreviadamente sería  $NID(0, \sigma^2)$ . Por lo que las observaciones  $y_i$  son  $NID(\beta_0 + \beta_1 x_i, \sigma^2)$ .

Supóngase que se desea hacer una prueba de hipótesis para cuando la pendiente es igual a una constante, la cual se llamará  $\beta_{10}$ . Estas serán:

$$H_0 : \beta_1 = \beta_{10} \text{ vs. } H_1 : \beta_1 \neq \beta_{10},$$

donde se tienen específicamente dos alternativas (que sea igual o no a  $\beta_{10}$ ). Ahora  $\hat{\beta}_1$  es una combinación lineal de las observaciones, entonces  $\hat{\beta}_1$  está normalmente distribuida con media  $\beta_1$  y varianza  $\sigma^2/S_{xx}$ . Por lo tanto el estadístico

$$Z_0 = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\sigma^2/S_{xx}}},$$

se distribuye *normal*(0,1) si la hipótesis nula  $H_0 : \beta_1 = \beta_{10}$  es verdadera. Si  $\sigma^2$  es conocida, se podría usar  $Z_0$  para la prueba antes mencionada. Sin embargo, los cuadrados medios de los residuos  $MS_E$  es un estimador insesgado de  $\sigma^2$ , y la distribución de  $(n-2)MS_E/\sigma^2$  es  $\chi_{n-2}^2$ . Además  $MS_E$  y  $\hat{\beta}_1$  son variables aleatorias



independientes. Estas condiciones implican que si se sustituye  $\sigma^2$  en  $Z_0$  por  $\hat{\sigma}^2 = MS_E$ , el estadístico

$$t_0 = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{MS_E/S_{xx}}},$$

se distribuye como una *t de Student* con  $(n-2)$  grados de libertad (que están asociados a  $MS_E$ ) si la hipótesis nula  $H_0 : \beta_1 = \beta_{10}$  es verdadera. El estadístico  $t_0$  es usado para la prueba  $H_0 : \beta_1 = \beta_{10}$  comparando los valores observados de  $t_0$  con un cuantil  $1 - \alpha/2$  de la distribución *t* con  $n-2$  grados de libertad. Se rechaza la hipótesis nula si

$$|t_0| > t_{n-2}^{1-\alpha/2}.$$

Un proceso similar puede ser utilizado para la prueba de hipótesis de  $\beta_0$ . Esta es

$$H_0 : \beta_0 = \beta_{00} \text{ vs } H_1 : \beta_0 \neq \beta_{00}.$$

Para ésta se utiliza el estadístico

$$t_0 = \frac{\hat{\beta}_0 - \beta_{00}}{\sqrt{MS_E(1/n + \bar{x}^2/S_{xx})}},$$

y se rechaza la hipótesis nula si  $|t_0| > t_{n-2}^{1-\alpha/2}$ .

Un caso especial es:

$$H_0 : \beta_1 = 0 \text{ vs } H_1 : \beta_1 \neq 0.$$

Esta hipótesis se refiere a la significancia de la regresión. No rechazar  $H_0 : \beta_1 = 0$  implica que no hay una relación lineal entre  $x$  y  $y$ .

Alternativamente si  $H_0 : \beta_1 = 0$  es rechazada, esto implica que  $x$  es de valor para explicar la variabilidad de  $y$ . Sin embargo, no aceptando  $H_0 : \beta_1 = 0$  podría significar que el modelo de línea recta es adecuado, o que incluso aunque exista un efecto lineal de  $x$  sobre  $y$ , mejores resultados podrían obtenerse con la adición de términos polinomiales de orden superior en  $x$ .

Otra utilidad de revisar la prueba de hipótesis es que se puede construir el estadístico de prueba  $F$ , el cual se resume en una tabla llamada análisis de varianza y que es conocida como la ANOVA (por sus siglas en ingles *Analysis of Variance*).

ANOVA				
<i>Fuente de variación</i>	<i>Suma de cuadrados</i>	<i>Grados de libertad</i>	<i>Cuadrados medios</i>	$F_0$
<b>Regresión</b>	$SS_R = \hat{\beta}_1 S_{xy}$	1	$MS_R$	$MS_R / MS_E$
<b>Residual</b>	$SS_E = S_{yy} - \hat{\beta}_1 S_{xy}$	$n - 2$	$MS_E$	
<b>Total</b>	$S_{yy}$	$n - 1$		

**Tabla 1.** Análisis de varianza para pruebas de significancia para la regresión

El análisis de varianza para la regresión consiste en descomponer la varianza total de la variable respuesta en varias partes llamadas fuentes de variación.

La división de la suma de cuadrados entre sus grados de libertad es llamada cuadrados medios, estos son:

- Cuadrado medio de la regresión  $MS_R = SS_R / 1$ .
- Cuadrado medio del error  $MS_E = SS_E / (n - 2)$ .

Si el estadístico calculado  $F_0$  es grande, entonces se tendrá un modelo que explica muy bien a los datos, ya que los cuadrados medios debidos al error serán pequeños.

## 1.6 Coeficiente de determinación y coeficiente de correlación

Una vez ajustada la recta de regresión a la nube de observaciones es importante disponer de una medida que calcule la bondad del ajuste realizado y que permita decidir si el ajuste lineal es suficiente o se deben buscar modelos alternativos. Como medida de bondad del ajuste se utiliza el coeficiente de determinación, definido como sigue

$$R^2 = \frac{SS_R}{S_{yy}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

o bien

$$R^2 = 1 - \frac{SS_E}{S_{yy}} = 1 - \left( \frac{n-2}{n-1} \right) \left( \frac{MS_E}{\hat{S}_{yy}^2} \right).$$

Como  $SS_R \leq S_{yy}$ , se verifica que  $0 \leq R^2 \leq 1$ .

El coeficiente de determinación mide la proporción de variabilidad total de la variable dependiente  $y$  respecto a su media que es explicada por el modelo de regresión. Es usual expresar esta medida en términos porcentuales, multiplicándola por cien.

Por otra parte, teniendo en cuenta que  $\hat{y}_i - \bar{y} = \hat{\beta}_1(x_i - \bar{x})$ , se obtiene

$$R^2 = \frac{S_{XY}^2}{S_X^2 S_Y^2}.$$

Dadas dos variables aleatorias cualesquiera  $X$  y  $Y$  una medida de la relación lineal que hay entre ambas variables es el llamado coeficiente de correlación, el cual está definido por

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}.$$

Un buen estimador de este parámetro es el coeficiente de correlación lineal muestral (o coeficiente de correlación de Pearson), definido por

$$r = \frac{S_{XY}}{S_X S_Y} = \text{signo}(\hat{\beta}_1) \sqrt{R^2}, \quad r \in [-1, 1],$$

donde *signo* es el signo que acompaña a  $\hat{\beta}_1$ .

Este coeficiente es una buena medida de la bondad del ajuste de la recta de regresión. Evidentemente, existe una estrecha relación entre  $r$  y  $\hat{\beta}_1$  aunque estos estimadores proporcionan diferentes interpretaciones del modelo:

- $r$  es una medida de la relación lineal entre las variables  $X$  y  $Y$ .
- $\hat{\beta}_1$  mide el cambio producido en la variable  $y$  al realizarse un incremento de una unidad en la variable  $x$ .

De las definiciones anteriores se deduce que:

$$S_{XY} = 0 \Leftrightarrow \hat{\beta}_1 = 0 \Leftrightarrow r = 0.$$

Es importante estudiar si  $r$  es significativo (distinto de cero) ya que ello implica que el modelo de regresión lineal será representativo de los datos. Desafortunadamente la distribución de  $r$  es complicada pero para tamaños muestrales mayores a 30 observaciones su desviación típica es  $\sqrt{\text{Var}(r)} \approx 1/\sqrt{n}$ , y puede utilizarse la siguiente regla

$$|r| > \frac{2}{\sqrt{n}} \Rightarrow r \text{ es significativo (con } \alpha = 0.05 \text{)}.$$

### 1.6.1 Interpretación del coeficiente de correlación.

Como se ha indicado el coeficiente de correlación de Pearson es un índice cuyos valores absolutos oscilan entre 0 y 1. Cuanto más cerca se esté del 1 mayor será la correlación, y menor cuanto más cerca se esté del cero. Pero, ¿cómo interpretar a este coeficiente?, ¿qué significa un coeficiente de 0.6?, ¿es alto o bajo? No puede darse una respuesta precisa. Depende en gran parte de la naturaleza de la investigación. Por ejemplo, una correlación de 0.6 sería baja si se trata de la fiabilidad de un cierto *test*, pero sin embargo, sería alta si se está hablando de su validez.

No obstante, se intentará abordar el tema desde dos perspectivas distintas. Por un lado, ya ha sido tratado desde el criterio de la significación estadística señalada en el apartado anterior. Desde este enfoque una correlación es efectiva si puede afirmarse que es distinta de cero, pero ha de decirse que para ser significativa no necesariamente ha de ser elevada; simplemente es diferente de cero. O en otros términos, es una correspondencia que es poco probable que proceda de una población cuya correlación es cero. Tan solo se está diciendo que se ha obtenido "algo" y que ese "algo" es (probablemente) más que "nada". La significación de  $r$  depende en gran medida del tamaño de la muestra, por ejemplo, una correlación de 0.01 puede ser significativa en una muestra suficientemente grande y otra de 0.9 no serlo en una pequeña. Aquí se cumple la ley de los grandes números; tendencias débiles son muy improbables, desde la hipótesis nula, en mayores masas de datos, mientras que tendencias fuertes pueden ser relativamente probables en un tamaño pequeño de muestra.

Por lo tanto, para la interpretación del coeficiente de correlación se debe tener en cuenta que:

- $r = \pm 1$  indica una relación lineal exacta positiva (creciente) o negativa (decreciente).
- $r = 0$  indica la no existencia de relación lineal estocástica, pero no indica independencia de las variables ya que puede existir una relación no lineal incluso exacta.

- valores intermedios de  $r$  ( $0 < r < 1$  ó  $-1 < r < 0$ ) indican la existencia de una relación lineal estocástica, más fuerte cuanto más próximo a +1 (ó -1) sea el valor de  $r$ .

Más interés tiene la interpretación del coeficiente de correlación en términos de la proporción de variabilidad compartida o explicada, donde se ofrece una idea más cabal de la magnitud de la relación. Esta explicación es proporcionada por el coeficiente de determinación. Éste se define como el cuadrado del coeficiente de correlación; esto es, dadas dos variables  $x$  y  $y$ , la  $R^2$  se entiende como una proporción de variabilidades. Por ejemplo, si la correlación entre inteligencia y rendimiento académico es de 0.8, significa que  $0.8^2 = 0.64$  es la proporción de varianza compartida entre ambas variables. Puede interpretarse diciendo que un 64% del rendimiento académico es debido a la inteligencia (variabilidad explicada), y esto es más exacto si se ha de ser estricto, que inteligencia y rendimiento académico comparten un 64% de elementos, o dicho en otras palabras, tanto la inteligencia como el rendimiento ponen en juego un 64% de habilidades comunes.

### 1.6.2 El $R^2$ ajustado

Para subsanar la tendencia del coeficiente de determinación, se ha definido un  $R^2$  ajustado de la siguiente manera:

$$R_{ajus}^2 = 1 - \frac{SS_E / (n - m)}{S_{yy} / (n - 1)} = 1 - \frac{n - 1}{n - m} (1 - R^2)$$

donde,  $m$  es el número de parámetros en el modelo.

El modelo que se busca es aquel que tiene un  $R^2$  ajustado alto.

En palabras de Mallows (1973), “el mejor modelo es aquél que no tiene ni mucha falta de ajuste ni mucho sobre ajuste para aproximar los datos”.

- Falta de ajuste, se da cuando el estimado del valor predicho de la variable de respuesta tiene mucho sesgo y poca varianza.
- Sobre ajuste, se da cuando la varianza del estimado del valor predicho es bastante alta, pero el sesgo es bajo.

El  $R^2$  ajustado se utiliza principalmente en los modelos multivariados.

## 1.7 Predicción en regresión lineal simple

Hay dos intenciones principales en el ajuste de un modelo de regresión:

- Explicar la relación existente entre la variable independiente y la dependiente. En el caso de la regresión lineal simple se estima la mejor recta que relaciona las variables  $y$  y  $x$ , para que posteriormente se cuantifique la importancia de dicha asociación por medio del coeficiente de correlación.
- Utilizar el modelo de regresión ajustado para predecir el valor de la variable respuesta  $y$  cuando la explicativa toma un valor determinado,  $x = x_j$ , para alguna  $j \in \{1, 2, \dots, n\}$ , donde  $n$  es el número de observaciones de la muestra.

En esta sección se estudia este segundo tema. Por lo que ya estimada la recta de regresión, ¿cómo poder predecir el valor de  $y$  sabiendo que  $x = x_j$ ? Esto es, pronosticar un valor de la variable condicionada  $y|x = x_j$ . Con lo que se quiere responder a preguntas del tipo: “La empresa MEGA tiene ingresos anuales de 300 millones, ¿cuál será el gasto en material informático de esta empresa?”.

### 1.7.1 Predicción de nuevas observaciones

Una importante aplicación del modelo de regresión es la predicción de nuevas observaciones  $y$  correspondientes a un valor específico de  $x$ . Si  $x_0$  es el valor de la variable independiente, entonces

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

es el punto estimado del nuevo valor de  $y_0$ .

Ahora se puede obtener un intervalo de predicción para las observaciones futuras  $y_0$ .

Habrá que notar que la variable aleatoria

$$\psi = y_0 - \hat{y}_0$$

está distribuida normalmente con media cero y varianza

$$\begin{aligned} V(\psi) &= V(y_0 - \hat{y}_0) \\ &= \sigma^2 \left[ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right], \end{aligned}$$

esto es porque la observación futura  $y_0$  es independiente de  $\hat{y}_0$ . Si se utiliza  $\hat{y}_0$  para predecir a  $y_0$ , entonces el error estándar de  $\psi = y_0 - \hat{y}_0$  es el estadístico apropiado sobre el cual se basa un intervalo de predicción. Así el intervalo al  $100(1 - \alpha)\%$  para la predicción sobre las observaciones futuras en  $x_0$  es

$$\hat{y}_0 - t_{n-2}^{1-\alpha/2} \sqrt{MS_E \left( 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \leq y_0 \leq \hat{y}_0 + t_{n-2}^{1-\alpha/2} \sqrt{MS_E \left( 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}.$$

El intervalo de predicción anterior es de anchura mínima en  $x_0 = \bar{x}$  y se amplía cuando  $|x_0 - \bar{x}| > 0$ .

Ahora, si se comparan los dos últimos intervalos de confianza y de predicción, los cuales son:



intervalo de confianza

$$\hat{y}_0 - t_{n-2}^{1-\alpha/2} \sqrt{MS_E \left( \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \leq E(y | x_0) \leq \hat{y}_0 + t_{n-2}^{1-\alpha/2} \sqrt{MS_E \left( \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)},$$

e

intervalo de predicción

$$\hat{y}_0 - t_{n-2}^{1-\alpha/2} \sqrt{MS_E \left( 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \leq y_0 \leq \hat{y}_0 + t_{n-2}^{1-\alpha/2} \sqrt{MS_E \left( 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)},$$

se puede observar que el intervalo de predicción en  $x_0$  es siempre más ancho que el de confianza porque el primero de ellos depende del error del modelo ajustado y del error asociado con las futuras observaciones.

## 2da Parte: Regresión bayesiana

### 2.1 Teoría de la decisión

En gran parte de los fenómenos físicos, biológicos y sociales se desconocen las leyes que regulan sus comportamientos, por lo que no es arriesgado pensar que estén gobernados por elementos de naturaleza estadística.

Bajo esta premisa, a todo decisor se le plantea el problema de elegir entre las acciones alternativas posibles en cada caso, sin tener el conocimiento preciso para optar con certeza por la acción más adecuada, basándose en la mayor o menor especificación que realice sobre los elementos de naturaleza estadística que han de remplazar a las leyes deterministas desconocidas sobre dicho fenómeno.

En general, un elemento estadístico estará representado por la distribución de probabilidad  $F(x; \theta)$  del fenómeno analizado, donde para cada posible valor del parámetro  $\theta$ , se obtiene uno u otro comportamiento concreto del fenómeno analizado.

intervalo de confianza

$$\hat{y}_0 - t_{n-2}^{1-\alpha/2} \sqrt{MS_E \left( \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \leq E(y | x_0) \leq \hat{y}_0 + t_{n-2}^{1-\alpha/2} \sqrt{MS_E \left( \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)},$$

e

intervalo de predicción

$$\hat{y}_0 - t_{n-2}^{1-\alpha/2} \sqrt{MS_E \left( 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \leq y_0 \leq \hat{y}_0 + t_{n-2}^{1-\alpha/2} \sqrt{MS_E \left( 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)},$$

se puede observar que el intervalo de predicción en  $x_0$  es siempre más ancho que el de confianza porque el primero de ellos depende del error del modelo ajustado y del error asociado con las futuras observaciones.

## 2da Parte: Regresión bayesiana

### 2.1 Teoría de la decisión

En gran parte de los fenómenos físicos, biológicos y sociales se desconocen las leyes que regulan sus comportamientos, por lo que no es arriesgado pensar que estén gobernados por elementos de naturaleza estadística.

Bajo esta premisa, a todo decisor se le plantea el problema de elegir entre las acciones alternativas posibles en cada caso, sin tener el conocimiento preciso para optar con certeza por la acción más adecuada, basándose en la mayor o menor especificación que realice sobre los elementos de naturaleza estadística que han de remplazar a las leyes deterministas desconocidas sobre dicho fenómeno.

En general, un elemento estadístico estará representado por la distribución de probabilidad  $F(x; \theta)$  del fenómeno analizado, donde para cada posible valor del parámetro  $\theta$ , se obtiene uno u otro comportamiento concreto del fenómeno analizado.

En la terminología de este enfoque, el conjunto de valores que puede tomar  $\theta$  recibe el nombre de estados de la naturaleza.

Brevemente se exponen los componentes fundamentales de un proceso de decisión:

- Conjunto de acciones, alternativas o estrategias que el decisor puede adoptar, se representa por  $A = \{a_1, \dots, a_h\}$ .
- Designar el espacio paramétrico como  $\Theta = \{\theta_1, \dots, \theta_k\}$ .
- Derivados de la elección de una determinada acción, según la cual sea el estado de la naturaleza  $\theta$ , se producen una serie de resultados que se representan por  $l(\theta; a)$ , donde cada valor será la pérdida en que se incurre cuando el decisor toma la acción  $a$  y la naturaleza se encuentra en el estado  $\theta$ . La expresión  $l(\theta; a)$  recibe el nombre de función de pérdida. Se dice que una pérdida negativa será igual a una ganancia.

Suponiendo que las acciones y los estados de la naturaleza son finitos, el proceso de decisión se recoge esquemáticamente en el cuadro siguiente:

		Estados de la naturaleza					
		$\theta_1$	$\theta_2$	•	•	•	$\theta_k$
Acciones	$a_1$	$l(\theta_1; a_1)$	$l(\theta_2; a_1)$	•	•	•	$l(\theta_k; a_1)$
	$a_2$	$l(\theta_1; a_2)$	$l(\theta_2; a_2)$	•	•	•	$l(\theta_k; a_2)$
	•	•	•	•	•	•	•
	•	•	•	•	•	•	•
	$a_h$	$l(\theta_1; a_h)$	$l(\theta_2; a_h)$	•	•	•	$l(\theta_k; a_h)$

**Cuadro 1**

## 2.2 Introducción a la estadística bayesiana

Afirmaciones del tipo “es muy probable que Juan haya sido quien llamó por teléfono” o “es probable que se encuentre un tratamiento eficaz para el SIDA en los próximos 5

años”, normales en el lenguaje común, no pueden cuantificarse formalmente; resultan ajenas, por tanto, a una metodología que se desenvuelva en un marco frecuentista. Una cuantificación sobre una base subjetiva resulta, sin embargo, familiar y enriquece el enfoque bayesiano<sup>2</sup>. Al admitir un manejo subjetivo de la probabilidad, el investigador bayesiano podrá emitir juicios de probabilidad sobre una hipótesis  $H$  y expresar por esa vía su grado de convicción al respecto, tanto antes como después de haber observado los datos. En su versión más elemental y en este contexto, el teorema de Bayes asume la forma siguiente:

$$P(H | \text{datos}) = [P(\text{datos} | H) / P(\text{datos})] \cdot P(H).$$

La probabilidad *a priori* de una hipótesis,  $P(H)$ , se ve transformada en una probabilidad *posteriori*,  $P(H | \text{datos})$ , una vez incorporada la evidencia que aportan los datos. El caso considerado se ajusta a la situación más simple, aquella en que  $P(H)$  representa un número único; sin embargo, si se consiguiera expresar la convicción inicial (y la incertidumbre) mediante una distribución de probabilidades entonces, una vez observados los datos, el teorema “devuelve” una nueva distribución, que no es otra cosa más que la percepción probabilística original actualizada por los datos.

Esta manera de razonar de la estadística bayesiana, es radicalmente diferente a la clásica, pero es sin embargo más cercana al modo de proceder cotidiano, ya que se trata de un proceso inductivo. Debe subrayarse que esta metodología, a diferencia del enfoque frecuentista, no tiene como finalidad producir una conclusión dicotómica (rechazo o aceptación), sino que cualquier información empírica, combinada con el conocimiento que ya se tenga del problema que se estudia, “actualiza” dicho conocimiento, y la trascendencia de dicha visión actualizada no depende de una regla mecánica.

---

<sup>2</sup> El adjetivo bayesiano proviene del uso frecuente que se hace del teorema de Bayes durante el proceso de inferencia. El teorema de Bayes se ha derivado del trabajo realizado por Thomas Bayes (matemático británico, nació en Londres, Inglaterra en 1702 y murió en Tunbridge Wells en 1761).

## 2.3 Inferencia bayesiana: distribuciones *a priori* y *posteriori*

La inferencia bayesiana está basada en la distribución de probabilidad del parámetro dado los datos, en lugar de la distribución de los datos dado el parámetro. Lo único que se requiere para realizar este proceso, es la especificación previa de una distribución de probabilidad, la cual representa el conocimiento acerca del parámetro antes de obtener cualquier información respecto a los datos.

Para la estadística bayesiana el concepto de probabilidad se encuentra en el observador, siendo así una percepción subjetiva. Sin embargo, además de la muestra, también juega un papel fundamental la información previa o la historia que se posee relativa a los fenómenos que se tratan de modelizar.

El análisis bayesiano se puede dividir en las siguientes etapas:

1. **Elección de un modelo de probabilidad completo.** Elegir una distribución de probabilidad conjunta para todas las cantidades observables y no observables. El modelo debe ser consistente con el conocimiento acerca del problema fundamental y el proceso de recolección de la información.
2. **Condicionamiento de los datos observados.** Calcular e interpretar la distribución *posteriori* apropiada que se define como la distribución de probabilidad condicional de las cantidades no observadas de interés, dados los datos observados.
3. **Evaluación del ajuste del modelo y las implicaciones de la distribución *posteriori* resultante.** ¿Es el modelo apropiado a los datos? ¿qué tan sensibles son los resultados a las suposiciones de modelamiento de la primera etapa? Si fuese necesario, alterar o ampliar el modelo y repetir las tres etapas ya señaladas.

En el marco teórico en que se aplica la técnica bayesiana hay un parámetro poblacional respecto al cual se desea realizar inferencias y se tiene un modelo que determina la probabilidad de observar diferentes valores de  $x$ , bajo distintos valores de los parámetros. Sin embargo, la diferencia fundamental es que la inferencia bayesiana considera al parámetro desconocido  $\theta$  como una variable aleatoria y a la que por lo

tanto, se le asigna una determinada distribución de probabilidad, que recibe el nombre de distribución *a priori* y se representa por  $g(\theta)$ .

Desde el punto de vista probabilístico existen tres procedimientos fundamentales para establecer la distribución de probabilidad *a priori*: distribución no informativa, que corresponde a una ausencia de opinión o de conocimiento clínico *a priori* y por lo tanto no aporta información a lo que se observa en los datos. Distribución *a priori* escéptica, que considera que la probabilidad de que la hipótesis alternativa sea cierta (existe diferencia entre los grupos) es muy pequeña. Y por último, la distribución *a priori* entusiasta, que tiene razones fundadas para encontrar discrepancias por lo que determina que la probabilidad de que éstas sean cero en el grupo de interés tiene una probabilidad muy baja.

El problema radica en que la especificación y cuantificación de la distribución *a priori* no es una tarea sencilla, especialmente cuando se trata de modelos con más de un parámetro, como pueden ser los de regresión.

La información suministrada por una muestra aleatoria simple  $\underline{X}$ , bajo la óptica bayesiana, puede cambiar la idea del comportamiento probabilístico que se tuviera sobre el parámetro  $\theta$  y, en este sentido, es posible aceptar que exista una distribución *posteriori* de este parámetro, representada por  $g(\theta | \underline{X})$ , donde se recoge la modificación de la distribución de probabilidad de  $\theta$  cuando se dispone de la información muestral  $\underline{X}$ .

Si la distribución *a priori* del parámetro  $\theta$  se postula de tipo discreto, aplicando el teorema de Bayes pueden determinarse las correspondientes probabilidades *posteriori* para cada uno de los valores que tome el parámetro  $\theta$ . En efecto

$$P(\theta = \theta_i | \underline{X}) = \frac{P(\underline{X} | \theta = \theta_i)P(\theta = \theta_i)}{\sum_{i=1}^k P(\underline{X} | \theta = \theta_i)P(\theta = \theta_i)}.$$

Si la distribución *a priori*  $g(\theta)$  es continua, y la población de donde se extrae la muestra también, puede obtenerse la distribución *posteriori* del parámetro  $\theta$  teniendo en cuenta que:

- $f(\underline{X}, \theta)$  representa la función de densidad conjunta del vector aleatorio muestral  $\underline{X}$  y del parámetro  $\theta$ .
- $f(\underline{X} | \theta)$  es la función de densidad de la muestra  $\underline{X}$  para un  $\theta$  dado.
- $f(\underline{X})$  es la distribución marginal de la muestra  $\underline{X}$ .

Como se sabe

$$f(\underline{X} | \theta) = \frac{f(\underline{X}, \theta)}{g(\theta)},$$

$$g(\theta | \underline{X}) = \frac{f(\underline{X}, \theta)}{f(\underline{X})},$$

donde

$$f(\underline{X}, \theta) = f(\underline{X} | \theta)g(\theta),$$

$$f(\underline{X}, \theta) = g(\theta | \underline{X})f(\underline{X}),$$

por lo que

$$f(\underline{X} | \theta)g(\theta) = g(\theta | \underline{X})f(\underline{X})$$

y la distribución *posteriori* del parámetro  $\theta$  resulta

$$g(\theta | \underline{X}) = \frac{f(\underline{X} | \theta)g(\theta)}{f(\underline{X})},$$

obteniéndose la densidad marginal  $f(\underline{X})$  como

$$f(\underline{X}) = \int_{\Theta} f(\underline{X}; \theta) d\theta = \int_{\Theta} f(\underline{X} | \theta) g(\theta) d\theta$$

llegándose finalmente a que

$$g(\theta | \underline{X}) = \frac{f(\underline{X} | \theta) g(\theta)}{\int_{\Theta} f(\underline{X} | \theta) g(\theta) d\theta},$$

versión continua del teorema de Bayes.

## 2.4 Estimación de Bayes

El estimador de Bayes es la función de decisión, establecida sobre el espacio paramétrico  $\Theta$ , que produzca el menor riesgo esperado, al cual se le llamará  $B(d)$ .

El estimador será  $\hat{\theta} = d^*(\underline{X})$  tal que minimice la expresión

$$B(d) = \int_{\underline{X}} \int_{\Theta} l[\theta; d(\underline{X})] f(\underline{X} | \theta) g(\theta) d\theta d\underline{X}$$

Existe un procedimiento operativo para determinar la decisión de Bayes sin tener que describir todo el espacio muestral.

Como ya se había establecido la distribución *posteriori* del parámetro  $\theta$  es

$$g(\theta | \underline{X}) = \frac{f(\underline{X} | \theta) g(\theta)}{f(\underline{X})}$$

por consiguiente, el riesgo de Bayes queda



$$\begin{aligned}
B(d) &= \int_X \int_{\Theta} l[\theta; d(\underline{X})] f(\underline{X} | \theta) g(\theta) d\theta d\underline{X} \\
&= \int_X \int_{\Theta} l[\theta; d(\underline{X})] g(\theta | \underline{X}) f(\underline{X}) d\theta d\underline{X} \\
&= \int_X f(\underline{X}) \left[ \int_{\Theta} l[\theta; d(\underline{X})] g(\theta | \underline{X}) d\theta \right] d\underline{X} \\
&= \int_X f(\underline{X}) E_{\Theta|\underline{X}}[l(\theta; d(\underline{X}))] d\underline{X}
\end{aligned}$$

siendo  $E_{\Theta|\underline{X}}[l(\theta; d(\underline{X}))]$ , e l riesgo de Bayes *posteriori*, es decir,

$$E_{\Theta|\underline{X}}[l(\theta; d(\underline{X}))] = \int_{\Theta} l[\theta; d(\underline{X})] g(\theta | \underline{X}) d\theta.$$

Dado que la función de decisión  $d(\underline{X})$  que minimiza a  $B(d)$  es la misma que hace mínima la expresión anterior, puede utilizarse esta última para determinar el estimador de Bayes, de manera que dicho estimador será aquel que verifique

$$\frac{\partial E_{\Theta|\underline{X}}[l(\theta; d(\underline{X}))]}{\partial d(\underline{X})} = \psi[\theta; d(\underline{X})] = 0 \Rightarrow \theta^* = d^*(\underline{X}),$$

$$\left[ \frac{\partial^2 E_{\Theta|\underline{X}}[l(\theta; d(\underline{X}))]}{\partial d^2(\underline{X})} \right]_{d(\underline{X})=d^*(\underline{X})} > 0.$$

## 2.5 Familias conjugadas

Tanto  $g(\theta)$  como  $g(\theta | \underline{X})$  son distribuciones de probabilidad sobre  $\theta$ , la primera sólo incorpora información *a priori* y la segunda actualiza dicha información junto con la que pueda adquirirse de la muestra. Si bien, la elección de alguna distribución de probabilidad para modelar la incertidumbre sobre  $\theta$  no resulta crucial en tanto sea factible obtener con cualquiera de ellas una distribución *a priori*, resulta conveniente

tanto para el análisis como desde un punto de vista computacional el que  $g(\theta)$  y  $g(\theta | \underline{X})$  pertenezcan a la misma familia. La definición siguiente resume lo mencionado anteriormente. Sea  $P = \{g(\underline{X} | \theta) : \theta \in \Theta\}$  una familia paramétrica. Una colección de distribuciones de probabilidad  $F$  es una familia conjugada para  $P$  si para todo  $g(\underline{X} | \theta) \in P$  y  $g(\theta) \in F$  se cumple que  $g(\theta | \underline{X}) \in F$ .

## 2.6 Distribuciones *a priori* no informativas

La estadística bayesiana proporciona una metodología que permite combinar de manera consistente información *a priori* con información muestral. Ante esto surge la pregunta de cómo realizar inferencias cuando no se dispone de información *a priori*, o bien cuando dicha información no se quiere o no se puede utilizar.

El problema quedaría resuelto si pudiésemos determinar una distribución *a priori* que describa la situación en que los datos experimentales contienen toda la información relevante, en lugar de proporcionar tan sólo parte de ella como sucede cuando se dispone de información *a priori*. Una forma pragmática (pero incompleta) de atacar este problema sería asignar arbitrariamente una distribución *a priori* con la única condición de que tenga una varianza “muy grande”.

Otra forma sería la siguiente. Supóngase que se tiene un número finito de sucesos inciertos (o hipótesis)  $E_1, \dots, E_n$ . Una distribución *a priori* que describe un estado de ignorancia o carencia de información es la siguiente:

$$U(E_j) = \frac{1}{n} \mathbf{I}_{\{1, \dots, n\}}(j)$$

esto es, una distribución uniforme discreta. Thomas Bayes propuso esta distribución con base en lo que él llamó el principio de la razón insuficiente: si no se sabe cosa alguna sobre  $\{E_1, \dots, E_n\}$  no hay razón para asignarle a alguno de los sucesos inciertos una probabilidad diferente que a los otros. Sin embargo, este principio no es aplicable en situaciones donde el número de sucesos inciertos no es finito.

## 2.7 Regla de Jeffreys

Sólo en el caso en que el espacio paramétrico  $\Theta$  sea finito el principio de la razón insuficiente provee distribuciones *a priori* no informativas que son invariantes ante transformaciones uno-a-uno del parámetro (o vector de parámetros). Esto es poco útil cuando se tratan con espacios paramétricos infinitos.

Jeffreys (1961) propuso una clase de distribuciones *a priori* no informativas para el caso de espacios paramétricos infinitos. En términos generales, la construcción de esta clase consiste en buscar simultáneamente invarianzas ante transformaciones y proveer la menor información *a priori* en relación a la información muestral, vía la información de Fisher.

## 2.8 Modelo bayesiano de regresión lineal simple

Para el método bayesiano se aplica el mismo principio de regresión que en la estadística clásica (ver sección 1.3), es decir, se puede utilizar el método de mínimos cuadrados para estimar los parámetros  $\beta_0$  y  $\beta_1$  (ver sección 1.4.2), además de  $\sigma^2$  (ver sección 1.4.3). De igual manera se harán suposiciones sobre los errores  $\varepsilon_i$ , así como para la esperanza (ver sección 1.3).

La notación para esta sección será distinta de la utilizada en el primer capítulo.

### 2.8.1 Regresión por mínimos cuadrados

“Supóngase que se tiene un conjunto de buenos modelos. La idea se basa en asignar probabilidades *a priori* a los coeficientes de cada uno de estos modelos que incluyen sólo un subconjunto de predictoras e igualmente se asignan probabilidades *a priori* a

cada uno de los modelos. Al final se elegirá como mejor modelo aquel que tiene la probabilidad posterior más alta con respecto a la variable de respuesta.”<sup>3</sup>

Se puede dibujar cualquier número de rectas sobre la nube de puntos, alguna de ellas podría ajustarse bastante bien a estos, pero otras podrían estar extremadamente lejos. Un residual es la distancia vertical desde un punto observado sobre otro que está en la recta. El método de mínimos cuadrados puede ser utilizado para encontrar la mejor recta que se ajuste a los puntos en términos de poder minimizar lo más posible la suma de los cuadrados de los residuos.

La ecuación de una recta se determina por dos cosas: la pendiente  $\beta_1$  y su intersección con el eje de las  $y$ 's que se le llamará  $\alpha_0$ . En realidad, el primero de estos parámetros y cualquier otro punto sobre la recta lo hará, por ejemplo,  $\alpha_{\bar{x}}$  la intersección de la línea vertical en  $\bar{x}$ . Encontrar la recta por mínimos cuadrados es equivalente a hallar su pendiente y la ordenada al origen.

## 2.8.2 Forma alternativa de la recta de mínimos cuadrados

La pendiente y cualquier otro punto además de la ordenada al origen determinan la recta de regresión. Es decir, el punto  $A_{\bar{x}}$ , donde la recta por mínimos cuadrados interseca al eje vertical en  $\bar{x}$ :

$$A_{\bar{x}} = A_0 + B\bar{x} = \bar{y}.$$

Por lo que la recta de mínimos cuadrados pasa a través del punto  $(\bar{x}, \bar{y})$ . Una ecuación alternativa para esta recta es

$$y = A_{\bar{x}} + B(x - \bar{x}) = \bar{y} + B(x - \bar{x})$$

que es especialmente útil.

---

<sup>3</sup> Mitchel and Beauchamp (1988): *Bayesian variable selection in linear regression*. JASSA

### 2.8.3 Estimación de la varianza por mínimos cuadrados

La estimación de la varianza para la recta de regresión por el método de mínimos cuadrados es

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n [y_i - (A_x - B(x_i - \bar{x}))]^2}{n - 2}$$

que es la suma de los cuadrados de los residuos dividido por  $n - 2$ . La razón de que sea  $n - 2$  es que se están utilizando dos estimadores,  $A_x$  y  $B$ , en el cálculo de la suma de los cuadrados<sup>4</sup>.

### 2.8.4 Supuestos de la recta de regresión

El método de mínimos cuadrados no hace uso de la distribución de probabilidad de los puntos. Esto es realmente una herramienta de análisis y puede ser aplicada a datos bivariados. No se puede hacer ninguna inferencia acerca de la pendiente y de la intersección ni sobre cualquier predicción del modelo de mínimos cuadrados, a menos que se realicen algunos supuestos acerca de modelo de probabilidad que determinan los datos. Los supuestos de la regresión son:

1. **Supuesto de la media.** La media condicional de  $y$  dado  $x$  es una función desconocida de  $x$ .

$$\mu_{y|x} = \alpha_0 + \beta_1 x$$

donde  $\beta_1$  es la pendiente desconocida y  $\alpha_0$  es la intersección con el eje de las  $y$ 's, la intersección con la línea vertical  $x = 0$ . Una parametrización alternativa será

---

<sup>4</sup> La regla general para encontrar un estimador insesgado de la varianza es que la suma de los cuadrados sea dividida entre los grados de libertad, y no hay que olvidar además que se pierde un grado por cada parámetro estimado en la fórmula de la suma de los cuadrados.

$$\mu_{y|x} = \alpha_{\bar{x}} + \beta_1(x - \bar{x})$$

donde  $\alpha_{\bar{x}}$  es la intersección desconocida con la línea vertical en  $x = \bar{x}$ . En esta parametrización los estimadores por mínimos cuadrados  $A_{\bar{x}} = \bar{y}$  y  $B$  serán independientes bajo los supuestos, entonces la *función de verosimilitud* será un factor en una parte dependiendo de  $\alpha_{\bar{x}}$  y en otra de  $\beta_1$ .

2. **Supuestos de los errores.** Estarán normalmente distribuidos con media cero y varianza constante  $\sigma^2$ .
3. **Supuestos de independencia.** Los errores para todas las observaciones serán independientes uno del otro.

Si se utiliza la parametrización alternativa

$$y_i = \alpha_{\bar{x}} + \beta_1(x - \bar{x}) + \varepsilon_i$$

donde  $\alpha_{\bar{x}}$  es el valor medio para  $y$  dado  $x = \bar{x}$ , y  $\beta_1$  es la pendiente. Como los  $\varepsilon_i$  son independientes entre sí, entonces  $y_i | x_i$  está normalmente distribuido con media  $\alpha_{\bar{x}} + \beta_1(x - \bar{x})$  y varianza constante  $\sigma^2$ , y por lo tanto  $y_i | x_i$  será independiente un de todos los demás.

## 2.9 Teorema de Bayes para el modelo de regresión

El teorema de Bayes se puede asumir como:

$$\text{función posteriori} \propto \text{función a priori} \times \text{función de verosimilitud}^5$$

---

<sup>5</sup>  $\propto$  es interpretado como el signo de una aproximación.

por lo que se necesitará determinar la *función de verosimilitud* y decidir la *a priori* para el modelo.

### 2.9.1 La *función de verosimilitud conjunta* para $\alpha_{\bar{x}}$ y $\beta_1$

La *función de verosimilitud* conjunta de la *i*-ésima observación se ve como una función de densidad de probabilidad de dos parámetros  $\alpha_{\bar{x}}$  y  $\beta_1$ , donde  $(x_i, y_i)$  son ajustados en los puntos observados. Esto da un peso relativo a todos los valores posibles de ambos parámetros  $\alpha_{\bar{x}}$  y  $\beta_1$  de las observaciones. La *función de verosimilitud* de la *i*-ésima observación es

$$\text{función de verosimilitud}(\alpha_{\bar{x}}, \beta_1) \propto e^{-\frac{1}{2\sigma^2}[y_i - (\alpha_{\bar{x}} + \beta_1(x_i - \bar{x}))]^2}$$

ya que se puede ignorar la parte que no contenga a los parámetros. Las observaciones son independientes, entonces la *función de verosimilitud* del total de la muestra de todas las observaciones es el producto individual de las *funciones de verosimilitud*:

$$\text{función de verosimilitud}_{\text{muestra}}(\alpha_{\bar{x}}, \beta_1) \propto \prod_{i=1}^n e^{-\frac{1}{2\sigma^2}[y_i - (\alpha_{\bar{x}} + \beta_1(x_i - \bar{x}))]^2}$$

pero para el producto de exponenciales solamente se suman los exponentes, entonces

$$\text{función de verosimilitud}_{\text{muestra}}(\alpha_{\bar{x}}, \beta_1) \propto e^{-\frac{1}{2\sigma^2} \left[ \sum_{i=1}^n [y_i - (\alpha_{\bar{x}} + \beta_1(x_i - \bar{x}))]^2 \right]}.$$

Ahora se procederá con el término que está dentro de los corchetes

$$\left[ \sum_{i=1}^n [y_i - (\alpha_{\bar{x}} + \beta_1(x_i - \bar{x}))]^2 \right] = \left[ \sum_{i=1}^n [y_i - \bar{y} + \bar{y} - (\alpha_{\bar{x}} + \beta_1(x_i - \bar{x}))]^2 \right].$$

Elevando al cuadrado y separando en una suma de tres términos, se obtiene

$$\sum_{i=1}^n (y_i - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - \bar{y})(\bar{y} - (\alpha_x + \beta_1(x_i - \bar{x}))) + \sum_{i=1}^n (\bar{y} - (\alpha_x + \beta_1(x_i - \bar{x})))^2.$$

Simplificando

$$\sum_{i=1}^n (y_i - \bar{y})^2 - 2\beta_1 \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) + \beta_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 + n(\alpha_x - \bar{y})^2.$$

Reescribiéndola como

$$S_{yy} - 2\beta_1 S_{xy} + \beta_1^2 S_{xx} + n(\alpha_x - \bar{y})^2.$$

Así, la *función de verosimilitud* conjunta puede ser escrita como sigue

$$\text{función de verosimilitud}_{\text{muestra}}(\alpha_x, \beta_1) \propto e^{-\frac{1}{2\sigma^2}[S_{yy} - 2\beta_1 S_{xy} + \beta_1^2 S_{xx} + n(\alpha_x - \bar{y})^2]}.$$

Escribiendo esto como producto de dos exponenciales queda como

$$\text{función de verosimilitud}_{\text{muestra}}(\alpha_x, \beta_1) \propto e^{-\frac{1}{2\sigma^2}[S_{yy} - 2\beta_1 S_{xy} + \beta_1^2 S_{xx}]} e^{-\frac{1}{2\sigma^2/n}n(\alpha_x - \bar{y})^2}.$$

Tomando el primer factor  $S_{xx}$  de la primera exponencial, completando el cuadrado, y quitando la parte que no depende de ningún parámetro dentro de la constante de proporcionalidad, resulta

$$\text{función de verosimilitud}_{\text{muestra}}(\alpha_x, \beta_1) \propto e^{-\frac{1}{2\sigma^2/S_{xx}}\left[\beta_1 - \frac{S_{xy}}{S_{xx}}\right]^2} e^{-\frac{1}{2\sigma^2/n}n(\alpha_x - \bar{y})^2}.$$

Nótese que  $B = S_{xy}/S_{xx}$  y que  $\bar{y} = A_x$ , son la pendiente y la intersección estimadas por mínimos cuadrados, respectivamente, pero la segunda de estas en la línea vertical  $x = \bar{x}$ . Se tiene que factorizar la *función de verosimilitud* conjunta dentro del producto de dos funciones individuales



*función de verosimilitud*<sub>muestra</sub>  $(\alpha_{\bar{x}}, \beta_1) \propto$  *función de verosimilitud*<sub>muestra</sub>  $(\alpha_{\bar{x}}) \times$  *función de verosimilitud*<sub>muestra</sub>  $(\beta_1)$

donde

$$\text{función de verosimilitud}_{\text{muestra}}(\beta_1) \propto e^{-\frac{1}{2\sigma^2/S_{xx}}[\beta_1 - B]^2},$$

$$\text{función de verosimilitud}_{\text{muestra}}(\alpha_{\bar{x}}) \propto e^{-\frac{1}{2\sigma^2/n}[\alpha_{\bar{x}} - A_{\bar{x}}]^2}.$$

A parte de que la *función de verosimilitud* conjunta ha sido factorizada dentro del producto de las *funciones de verosimilitud* individuales, se sabe que éstas son independientes. Al mismo tiempo, la *función de verosimilitud* de la pendiente  $\beta_1$  tiene una forma normal con media  $B$  y varianza  $\sigma^2/S_{xx}$ . Similarmente la *función de verosimilitud* de  $\alpha_{\bar{x}}$  tiene forma normal con media  $A_{\bar{x}}$  y varianza  $\sigma^2/n$ .

### 2.9.2 La *a priori* conjunta para $\alpha_{\bar{x}}$ y $\beta_1$

Si se multiplica la *función de verosimilitud* conjunta por una *a priori* conjunta, esto será proporcional a la *posteriori* conjunta. Se utilizarán las *a priori* independientes para cada parámetro. La *a priori* conjunta de dos parámetros es el producto de dos *a priori* individuales:

$$g(\alpha_{\bar{x}}, \beta_1) = g(\alpha_{\bar{x}})g(\beta_1)$$

Se pueden utilizar las *funciones a priori* normales o las *a priori* no informativas, aunque se pueden elegir otras distribuciones *a priori*. El hecho de utilizar las normales es debido a la forma de la función conjugada (ver sección 2.5). Las no informativas pueden ser la de referencia, la de Jeffreys o cualquier otra (ver secciones 2.6 y 2.7)<sup>6</sup>.

<sup>6</sup> Se puede consultar el libro Bernardo, J.M. y Smith, A.F.M. (1994): *Bayesian Theory*. Chichester: Wiley.

### 2.9.3 Eligiendo las *funciones a priori* normales de $\alpha_x$ y $\beta_1$

Otra de las ventajas de utilizar esta parametrización es que se puede tener una mayor intuición *a priori* acerca de  $\alpha_x$  (la intersección de  $x = \bar{x}$ ), que acerca de  $\alpha_0$  (la intersección de  $y$  con el eje de las  $x$ 's). El investigador debe decidir cuál es el valor de la media de los valores de la  $y$ . Eso será denotado por  $m_{\alpha_x}$  y llamado media *a priori* para  $\alpha_x$ . Entonces habrá que pensar en los puntos por encima y por debajo de lo que uno considera como límites superior e inferior de los valores posibles de  $y$ . Hay que dividir la diferencia entre cuatro para obtener  $s_{\alpha_x}$ , que será la desviación estándar *a priori* de  $\alpha_x$ <sup>7</sup>, (a dicha diferencia se le conoce como rango de la regresión).

Usualmente se está más interesado en la pendiente  $\beta_1$ . En ocasiones, se requiere determinar si ésta puede ser cero, por lo que, se pudiera elegir  $m_{\beta_1} = 0$  como la media *a priori* para  $\beta_1$ . Entonces se debe de pensar en los límites superior e inferior del efecto de un incremento en  $x$  sobre una unidad de  $y$ . Si se divide la diferencia entre cuatro se obtendrá  $s_{\beta_1}$ , el cual se llamará desviación estándar *a priori* de  $\beta_1$ . En otros casos, se puede tener una creencia *a priori* acerca de la pendiente poseyendo datos previos. Se podría utilizar la *normal*( $m_{\beta_1}, (s_{\beta_1})^2$ ) que relaciona la creencia *a priori*.

### 2.9.4 La *función posteriori conjunta* para $\beta_0$ y $\beta_1$

La función *posteriori* conjunta es proporcional a la función *a priori* conjunta multiplicada por la *función de verosimilitud* conjunta:

$$g(\alpha_x, \beta_1 | \text{dato}) \propto g(\alpha_x, \beta_1) \times \text{función de verosimilitud}_{\text{muestra}}(\alpha_x, \beta_1),$$

donde el *dato* es el conjunto de pares ordenados  $(x_1, y_1), \dots, (x_n, y_n)$ . Tanto la *a priori* conjunta como la *función de verosimilitud* conjunta ambos factores están en una parte

---

<sup>7</sup> La diferencia dividida entre cuatro se considera como una representación de los cuartiles.

en función de  $\alpha_{\bar{x}}$  y en otra en función de  $\beta_1$ . Reordenando éstos se tiene el factor *posteriori* conjunta en las funciones marginales *posteriori*:

$$g(\alpha_{\bar{x}}, \beta_1 | \text{dato}) \propto g(\alpha_{\bar{x}} | \text{dato})g(\beta_1 | \text{dato}).$$

Ya que la *función posteriori* conjunta es el producto de las *posteriori* marginales, éstas son independientes. Cada una de estas *funciones posteriori* marginales pueden ser encontradas usando las reglas de simple actualización para las distribuciones normales, la cual se usa para las *a priori* normales y las no informativas. Por ejemplo, si se utiliza una *normal*( $m_{\beta_1}, (s_{\beta_1})^2$ ) *a priori* para  $\beta_1$ , se tendría una *normal*( $m'_{\beta_1}, (s'_{\beta_1})^2$ ) *posteriori*, donde

$$\frac{1}{(s'_{\beta_1})^2} = \frac{1}{s_{\beta_1}^2} + \frac{S_{xx}}{\sigma^2},$$

$$m'_{\beta_1} = \frac{\frac{1}{s_{\beta_1}^2}}{\frac{1}{(s'_{\beta_1})^2}} m_{\beta_1} + \frac{\frac{S_{xx}}{\sigma^2}}{\frac{1}{(s'_{\beta_1})^2}} B.$$

La *precisión* de la *función posteriori* es igual a la *precisión* de la *función a priori* más la *precisión* de la *función de verosimilitud*. La *media posteriori* es igual al promedio ponderado de la *media a priori* más la *media ponderada* de la *función de verosimilitud*, donde las ponderaciones son las proporciones de las *precisiones* de la *precisión posteriori*:

$$\text{precisión posteriori} = \text{precisión a priori} + \text{precisión función de verosimilitud},$$

donde

$$\text{precisión} = 1/\text{varianza}.$$

Similarmente, si se utiliza una  $normal(m_{\alpha_x}, s_{\alpha_x}^2)$  a priori para  $\alpha_x$ , se tendría  $normal(m'_{\alpha_x}, (s'_{\alpha_x})^2)$  posteriori, donde

$$\frac{1}{(s'_{\alpha_x})^2} = \frac{1}{s_{\alpha_x}^2} + \frac{n}{\sigma^2} \quad \text{y} \quad m'_{\alpha_x} = \frac{\frac{1}{s_{\alpha_x}^2}}{\frac{1}{(s'_{\alpha_x})^2}} m_{\alpha_x} + \frac{\frac{n}{\sigma^2}}{\frac{1}{(s'_{\alpha_x})^2}} A_x.$$

### 2.9.5 Intervalo de credibilidad bayesiano para la pendiente <sup>8</sup>

La distribución de  $\beta_1$  posteriori resume toda la creencia e información después de examinar los datos. Esto se puede resumir por un intervalo de credibilidad bayesiano para la pendiente al  $(1 - \alpha) \times 100\%$ . El cual será

$$m'_{\beta_1} \pm Z_{1-\alpha/2} \sqrt{(s'_{\beta_1})^2}.$$

Como no se conoce  $\sigma^2$  una aproximación sensible en este caso es, utilizar la estimación calculada de los residuales

$$\sigma^2 = \frac{\sum_{i=1}^n (y_i - (A_x - B(x_i - \bar{x})))^2}{n - 2}.$$

Se tiene que ampliar el intervalo de confianza para tener en cuenta el aumento de la incertidumbre debida a no conocer  $\sigma^2$ . Se hace esto utilizando una *t de Student* con  $n - 2$  grados de libertad en vez de valor normal estándar. Entonces el intervalo de credibilidad es

$$m'_{\beta_1} \pm t_{1-\alpha/2} \sqrt{(s'_{\beta_1})^2}$$

<sup>8</sup> También es conocido como intervalo de máxima densidad posterior, o por sus siglas en inglés HPDI (*Highest Posterior Density Interval*).

### 2.9.6 Intervalo de confianza frecuentista para la pendiente

Cuando la varianza  $\sigma^2$  es desconocida, el intervalo de confianza al  $(1-\alpha)\times 100\%$  para la pendiente es

$$B \pm t_{1-\alpha/2} \frac{\hat{\sigma}}{\sqrt{S_{xx}}}$$

donde  $\hat{\sigma}$  es el estimado de la desviación estándar calculada de los residuales de la recta de mínimos cuadrados. El intervalo de confianza tiene la misma forma que el de credibilidad bayesiano cuando se utiliza la distribución *a priori* no informativa para  $\beta_1$  y  $\alpha_x$ . Por supuesto la interpretación es diferente. Bajo los supuestos frecuentistas se está confiando un  $(1-\alpha)\times 100\%$  de que el intervalo contendrá al verdadero valor del parámetro, aunque se sabe que no se conoce dicho valor. Por lo que en este caso, el intervalo de confianza es equivalente al de credibilidad bayesiano.

### 2.9.7 Pruebas de hipótesis bayesianas

Si se quiere aplicar un proceso formal, la teoría de decisión bayesiana es el camino por recorrer, ya que es posible obtener una distribución de probabilidad sobre el parámetro de espacio, además una utilidad esperada puede hacer cálculos basados en los costos y beneficios de los diferentes resultados.

### 2.9.8 Prueba de hipótesis de una cola para la pendiente

A menudo se requiere determinar si el aumento de la cantidad de  $y$  asociado con un incremento de una unidad en  $x$  es mayor que algún valor  $\beta_0$ . Esto se puede hacer mediante la prueba

$$H_0 : \beta_1 \leq \beta_0 \text{ vs. } H_1 : \beta_1 > \beta_0$$

en el nivel de significancia  $\alpha$  en un sentido bayesiano. Al realizar la prueba con esta técnica, se calcula la probabilidad *posteriori* de los parámetros bajo la hipótesis nula. Esto es

$$P(\beta_1 \leq \beta_0 \mid \text{dato}) = \int_{-\infty}^{\beta_0} g(\beta_1 \mid \text{dato}) d\beta_1$$

$$= P\left(Z \leq \frac{\beta_0 - m'_{\beta_1}}{s'_{\beta_1}}\right).$$

Si esta probabilidad es menor que  $\alpha$ , entonces se rechaza  $H_0$  y se concluye que, efectivamente, la pendiente  $\beta_1$  es mayor que  $\beta_0$ .<sup>9</sup>

### 2.9.9 Prueba de hipótesis de dos colas para la pendiente

Si  $\beta_1 = 0$ , entonces la media de  $y$  no depende de  $x$ . Lo que esta prueba quiere es comparar

$$H_0 : \beta_1 = 0 \text{ vs. } H_1 : \beta_1 \neq 0$$

al nivel de significancia  $\alpha$  en un modo bayesiano, antes de utilizar el modelo de regresión para hacer predicciones. Para hacer esta prueba con esta técnica, habrá que ver en donde se encuentra el cero en relación con el intervalo de credibilidad. Si se está fuera de este, se rechaza  $H_0$ . De lo contrario, no se rechaza la hipótesis nula, y no se debe utilizar el modelo de regresión para hacer predicciones.

---

<sup>9</sup> Si se utiliza la estimación de la varianza, entonces se podría utilizar la distribución *t de student* con  $n - 2$  grados de libertad en vez de la normal estándar  $Z$ .

### 2.9.10 Similitudes entre pruebas de hipótesis frecuentista y bayesiana

Las estimaciones de máxima verosimilitud de los parámetros, de los errores estándar y las estimaciones bayesianas con distribución *a priori* no informativa, son equivalentes. En el enfoque bayesiano, una prueba de hipótesis es producida por la combinación de información actual de los datos junto con la recopilación de experiencia pasada.

Asintóticamente, los datos afectan la elección de la *a priori*, de modo que si había un conjunto infinito de datos, las *a priori* serían irrelevantes y los resultados bayesiano y frecuentista convergerían.

Las pruebas frecuentistas de una cola son básicamente equivalentes a lo que un bayesiano obtendría utilizando intervalos de credibilidad.

### 2.9.11 Diferencias entre pruebas de hipótesis frecuentista y bayesiana

La diferencia pragmática más importante entre prueba de hipótesis bayesiana y frecuentista, es que la primera de éstas técnicas son poco adecuadas para pruebas de dos colas, esto es debido a que la probabilidad de cero en una distribución continua es igual a cero. La mejor solución propuesta hasta ahora es calcular la probabilidad de que, por ejemplo, un coeficiente de regresión es en algunos rangos cercano a cero. Entonces, para dos colas el  $p\text{-value} = \text{Probabilidad} (-e < B < e)$ .

La otra diferencia importante es más filosófica. Los  $p\text{-value}$  frecuentistas violan el principio de máxima verosimilitud.

### 2.9.12 El principio de máxima verosimilitud

Si  $x$  y  $y$  son dos puntos de la muestra de tal manera que

$$L(\theta | x) = C(x, y)L(\theta | y)$$

entonces, las inferencias extraídas de  $x$  deben ser idénticas que para las de  $y$ <sup>10</sup>.

## 2.10 Distribución de la predicción para futuras observaciones

Hacer predicciones de futuras observaciones para valores específicos de  $x$  es uno de los propósitos más importantes del modelo de regresión. A menudo, después de haber establecido que en los datos existe una relación lineal entre la variable explicativa  $x$  y la de respuesta  $y$ , se puede utilizar esta asociación para hacer predicciones del siguiente valor  $y_{n+1} | x_{n+1}$ . Se podrían hacer mejores predicciones utilizando el valor de la variable explicativa que sin ella. El mejor pronóstico para  $y_{n+1}$  dado  $x_{n+1}$ , será

$$\bar{y}_{n+1} = \alpha_{\bar{x}} + \hat{\beta}_1(x_{n+1} - \bar{x})$$

donde  $\hat{\beta}_1$  es el estimador de la pendiente y  $\alpha_{\bar{x}}$  es el estimado de la intersección de la línea  $x = \bar{x}$ .

¿Qué tan buena será la predicción? Existen dos fuentes de incertidumbre. Primero, se está utilizando los valores estimados de los parámetros en la predicción, no los valores verdaderos, los cuales se desconocen. Se considera a los parámetros como variables aleatorias y a los que por lo tanto se les encontró su distribución *posteriori*. Segundo, la nueva observación  $y_{n+1}$  tiene su propio error de observación  $\varepsilon_{n+1}$ , que al igual que todos los anteriores es independiente a los demás. La distribución de la predicción de una nueva observación  $y_{n+1}$  dado el valor  $x_{n+1}$  y el dato para las dos fuentes de incertidumbre, que se denota como  $f(y_{n+1} | x_{n+1}, \text{dato})$ , es hallado por el teorema de Bayes.

---

<sup>10</sup>En inglés se conoce como *the likelihood principle*.



### 2.10.1 Distribución de la predicción

La distribución de la predicción se encuentra integrando sobre  $\alpha_x^-$  y  $\beta_1$  fuera de la función de distribución *posteriori* conjunta de la observación  $y_{n+1}$  y los parámetros dado el valor  $x_{n+1}$  y las observaciones previas del modelo,  $(x_1, y_1), \dots, (x_n, y_n)$ , a estas últimas se les conocen como el *dato*. Esto es

$$f(y_{n+1} | x_{n+1}, \text{dato}) = \int \int f(y_{n+1}, \alpha_x^-, \beta_1 | x_{n+1}, \text{dato}) d\alpha_x^- d\beta_1.$$

Integrando con respecto a los parámetros de la función *posteriori* conjunta, ésta es conocida como marginalización.

Se necesita determinar la distribución *posteriori* conjunta de los parámetros, la observación  $y_{n+1}$  dado el valor  $x_{n+1}$  y el *dato*:

$$f(y_{n+1}, \alpha_x^-, \beta_1 | x_{n+1}, \text{dato}) = f(y_{n+1} | \alpha_x^-, \beta_1, x_{n+1}, \text{dato}) g(\alpha_x^-, \beta_1 | x_{n+1}, \text{dato}).$$

La observación  $y_{n+1}$ , dado los parámetros  $\alpha_x^-$ ,  $\beta_1$  y el valor conocido de  $x_{n+1}$ , es al igual que las demás aleatoria en el modelo de regresión. Dados  $\alpha_x^-$  y  $\beta_1$ , todas las observaciones son independientes cada una de las demás. Esto significa que dados  $\alpha_x^-$  y  $\beta_1$ , la  $y_{n+1}$  no dependerá de los datos, que son las observaciones anteriores de la regresión. La *función posteriori* para  $\alpha_x^-$  y  $\beta_1$  se calculó a partir solamente del *dato* y no depende del siguiente valor de la predicción  $x_{n+1}$ . Entonces la distribución conjunta de una nueva observación, simplifica los parámetros como:

$$f(y_{n+1}, \alpha_x^-, \beta_1 | x_{n+1}, \text{dato}) = f(y_{n+1} | \alpha_x^-, \beta_1, x_{n+1}) g(\alpha_x^-, \beta_1 | \text{dato})$$

la cual es la distribución de la siguiente observación dados los parámetros por la función de distribución *posteriori* de los parámetros dado los datos anteriores. La observación  $y_{n+1}$  dado los parámetros, es decir,  $y_{n+1} | \alpha_x^-, \beta_1, x_{n+1}$  es aleatoria en el modelo de

regresión dado el valor  $x_{n+1}$ . Por las suposiciones hechas se sabe que está normalmente distribuida con media dada por una función lineal de los parámetros  $\mu_{n+1} = \alpha_{\bar{x}} + \beta_1(x_{n+1} - \bar{x})$  y varianza conocida  $\sigma^2$ .

Las distribuciones *posteriori* de los parámetros dados los datos anteriores, las cuales se encuentran utilizando las reglas planteadas en la sección anterior, son funciones independientes  $normal(m'_{\alpha_{\bar{x}}}, (s'_{\alpha_{\bar{x}}})^2)$  y  $normal(m'_{\beta_1}, (s'_{\beta_1})^2)$ , respectivamente. Ya que la próxima observación  $y_{n+1}$  sólo depende de los parámetros debido a la función lineal.

$$\mu_{n+1} = \alpha_{\bar{x}} + \beta_1(x_{n+1} - \bar{x}).$$

Se puede simplificar el problema al permitir que  $\mu_{n+1}$  sea sólo un parámetro. Los dos componentes  $\alpha_{\bar{x}}$  y  $\beta_1$  son independientes, por lo que la distribución *posteriori* de  $\mu_{n+1}$  será *normal* con media  $m'_{\mu} = m'_{\alpha_{\bar{x}}} + (x_{n+1} - \bar{x}) \times m'_{\beta_1}$  y varianza  $(s'_{\mu})^2 = (s'_{\alpha_{\bar{x}}})^2 + (x_{n+1} - \bar{x})^2 \times (s'_{\beta_1})^2$ .

Se encontrará la distribución de la predicción por la marginación de  $\mu_{n+1}$  fuera de la función *posteriori* conjunta de  $y_{n+1}$  y  $\mu_{n+1}$ :

$$\begin{aligned} f(y_{n+1} | x_{n+1}, \text{dato}) &= \int f(y_{n+1}, \mu_{n+1} | x_{n+1}, \text{dato}) d\mu_{n+1} \\ &= \int f(y_{n+1} | \mu_{n+1}, x_{n+1}, \text{dato}) g(\mu_{n+1} | x_{n+1}, \text{dato}) d\mu_{n+1} \\ &= \int f(y_{n+1} | \mu_{n+1}) g(\mu_{n+1} | x_{n+1}, \text{dato}) d\mu_{n+1} \\ &\propto \int e^{-\frac{1}{2\sigma^2}(y_{n+1} - \mu_{n+1})^2} e^{-\frac{1}{2(s'_{\mu})^2}(\mu_{n+1} - m'_{\mu})^2} d\mu_{n+1} \end{aligned}$$

$$\propto \int e^{-\frac{1}{2\sigma^2(s'_\mu)^2/(\sigma^2+(s'_\mu)^2)}\left(\mu_{n+1}-\frac{y_{n+1}(s'_\mu)^2+m'_\mu\sigma^2}{(s'_\mu)^2+\sigma^2}\right)^2} e^{-\frac{1}{2((s'_\mu)^2+\sigma^2)}(y_{n+1}-m'_\mu)^2} d\mu_{n+1}.$$

El segundo factor no depende de  $\mu_{n+1}$ , entonces puede ser sacado de la integral, por lo que queda

$$f(y_{n+1} | x_{n+1}, \text{dato}) \propto e^{-\frac{1}{2((s'_\mu)^2+\sigma^2)}(y_{n+1}-m'_\mu)^2}$$

Así se tiene que ésta es una  $normal(m'_y, (s'_y)^2)$ , donde  $m'_y = m'_\mu$  y  $(s'_y)^2 = (s'_\mu)^2 + \sigma^2$ . En consecuencia la media de la predicción de la siguiente observación  $y_{n+1}$  tomada en  $x_{n+1}$  es la media *posteriori* de  $\mu_{n+1} = \alpha_{\bar{x}} + \beta_1(x_{n+1} - \bar{x})$ , y la varianza de la predicción de  $y_{n+1}$  es la varianza *posteriori* de  $\mu_{n+1} = \alpha_{\bar{x}} + \beta_1(x_{n+1} - \bar{x})$  más la varianza de la observación  $\sigma^2$ .

### 2.10.2 Intervalo de credibilidad para la predicción

Muchas veces se desea encontrar un intervalo que tenga una probabilidad *posteriori* igual a  $1 - \alpha$  que contenga el siguiente valor  $y_{n+1}$ , el cual será observado en  $x_{n+1}$ . Esto se conocerá como intervalo de credibilidad al  $(1 - \alpha) \times 100\%$  para la predicción. Se sabe que la media y la varianza de la distribución de predicción son  $m'_y$  y  $(s'_y)^2$ , respectivamente. Por lo que este es

$$\begin{aligned} m'_y \pm z_{\frac{\alpha}{2}} s'_y &= m'_\mu \pm z_{\frac{\alpha}{2}} \sqrt{(s'_\mu)^2 + \sigma^2} \\ &= m'_{\alpha_{\bar{x}}} + m'_{\beta_1} (x_{n+1} - \bar{x}) \pm z_{\frac{\alpha}{2}} \sqrt{(s'_{\alpha_{\bar{x}}})^2 + (s'_{\beta_1})^2 (x_{n+1} - \bar{x})^2 + \sigma^2} \end{aligned}$$

cuando se conoce la varianza de la observación  $\sigma^2$ .

Cuando no se conoce la varianza de la observación y se utiliza la estimación de la misma calculada de los residuales, el intervalo de credibilidad será

$$m'_y \pm t_{\frac{\alpha}{2}} s'_y = m'_\mu \pm t_{\frac{\alpha}{2}} \sqrt{(s'_\mu)^2 + \sigma^2}$$

$$= m'_{\alpha_x} + m'_{\beta_1} (x_{n+1} - \bar{x}) \pm t_{\frac{\alpha}{2}} \sqrt{(s'_{\alpha_x})^2 + (s'_{\beta_1})^2 (x_{n+1} - \bar{x})^2 + \sigma^2}$$

donde se obtiene el valor de la distribución *t de Student* con  $n - 2$  grados de libertad. Este intervalo de credibilidad para la predicción son los análogos bayesianos de los de predicción frecuentistas, ya que ellos tienen en cuenta el error de la estimación y el de observación. El primero de los intervalos generalmente es más chico que el segundo ya que los bayesianos usan información *a priori* así como información del *dato*. Estos dan exactamente el mismo resultado que el intervalo frecuentista de la predicción cuando las *a priori* no informativas son utilizadas tanto para la pendiente como para la intersección.

### 3ra Parte: Ejemplos de regresión lineal múltiple con enfoque clásico

Se eligieron tres problemas para desarrollarlos con las técnicas revisadas en los dos capítulos anteriores.

El programa que se utilizará para la aplicación de estos ejemplos es conocido como **Statistica** en su **versión 8**, pues éste maneja un software para el análisis de regresión múltiple clásico, por el método de mínimos cuadrados. Se comentará el funcionamiento de dicho programa para poder utilizar los menús y realizar los análisis de regresión.

#### 3.1 Ejemplo 1

El primero de los problemas se tomó del programa **Statistica**. Este ejemplo se llama *Poverty.sta* y se encuentra en el archivo de datos de dicho programa el cual se abre a

Cuando no se conoce la varianza de la observación y se utiliza la estimación de la misma calculada de los residuales, el intervalo de credibilidad será

$$m'_y \pm t_{\frac{\alpha}{2}} s'_y = m'_\mu \pm t_{\frac{\alpha}{2}} \sqrt{(s'_\mu)^2 + \sigma^2}$$

$$= m'_{\alpha_x} + m'_{\beta_1} (x_{n+1} - \bar{x}) \pm t_{\frac{\alpha}{2}} \sqrt{(s'_{\alpha_x})^2 + (s'_{\beta_1})^2 (x_{n+1} - \bar{x})^2 + \sigma^2}$$

donde se obtiene el valor de la distribución *t de Student* con  $n - 2$  grados de libertad. Este intervalo de credibilidad para la predicción son los análogos bayesianos de los de predicción frecuentistas, ya que ellos tienen en cuenta el error de la estimación y el de observación. El primero de los intervalos generalmente es más chico que el segundo ya que los bayesianos usan información *a priori* así como información del *dato*. Estos dan exactamente el mismo resultado que el intervalo frecuentista de la predicción cuando las *a priori* no informativas son utilizadas tanto para la pendiente como para la intersección.

### 3ra Parte: Ejemplos de regresión lineal múltiple con enfoque clásico

Se eligieron tres problemas para desarrollarlos con las técnicas revisadas en los dos capítulos anteriores.

El programa que se utilizará para la aplicación de estos ejemplos es conocido como **Statistica** en su **versión 8**, pues éste maneja un software para el análisis de regresión múltiple clásico, por el método de mínimos cuadrados. Se comentará el funcionamiento de dicho programa para poder utilizar los menús y realizar los análisis de regresión.

#### 3.1 Ejemplo 1

El primero de los problemas se tomó del programa **Statistica**. Este ejemplo se llama *Poverty.sta* y se encuentra en el archivo de datos de dicho programa el cual se abre a

través de *File – Open Examples*, que se encuentra en la carpeta llamada *Datasets*. Los datos se basan en una comparación de los Censos 1960 y 1970, son las cifras de una muestra aleatoria de 30 condados de Estados Unidos.

Los datos tomados son los que se muestran en la siguiente tabla.

Tabla 3.1.1

	Predictors of poverty						
	1	2	3	4	5	6	7
	POP_CHNG	N_EMPLD	PT_POOR	TAX_RATE	PT_PHONE	PT_RURAL	AGE
Benton	13.7	400	19.0	1.09	82	74.8	33.5
Cannon	-0.8	710	26.2	1.01	66	100.0	32.8
Carrol	9.6	1610	18.1	0.40	80	69.7	33.4
Cheatheam	40.0	500	15.4	0.93	74	100.0	27.8
Cumberland	8.4	640	29.0	0.92	65	74.0	27.9
DeKalb	3.5	920	21.6	0.59	64	73.1	33.2
Dyer	3.0	1890	21.9	0.63	82	52.3	30.8
Gibson	7.1	3040	18.9	0.49	85	49.6	32.4
Greene	13.0	2730	21.1	0.71	78	71.2	29.2
Hawkins	10.7	1850	23.8	0.93	74	70.6	28.7
Haywood	-16.2	2920	40.5	0.51	69	64.2	25.1
Henry	6.6	1070	21.6	0.80	85	58.3	35.9
Houston	21.9	160	25.4	0.74	69	100.0	31.4
Humphreys	17.8	380	19.7	0.44	83	72.0	30.1
Jackson	-11.8	1140	38.0	0.81	54	100.0	34.1
Johnson	7.5	690	30.1	1.05	65	100.0	30.5
Lawrence	3.7	1170	24.8	0.73	76	69.5	30.0
McNairy	1.6	1280	30.3	0.65	67	81.0	32.4
Madison	8.4	2270	19.5	0.48	85	39.1	28.7
Marshall	2.7	960	15.6	0.72	84	58.4	33.4
Maury	5.6	1710	17.2	0.62	84	42.4	29.9
Montgomery	12.7	1410	18.4	0.84	86	36.4	23.3
Morgan	-4.8	200	27.3	0.73	66	99.8	27.5
Sevier	16.5	960	19.2	0.45	74	90.6	29.5
Shelby	15.2	11500	16.8	1.00	87	5.9	25.4
Sullivan	11.6	1380	13.2	0.63	85	44.2	28.8
Trousdale	4.9	530	29.7	0.54	70	100.0	33.1
Unicoi	1.1	370	19.8	0.98	75	52.6	30.8
Wayne	3.8	440	27.7	0.46	48	100.0	28.4
Weakley	19.0	1630	20.5	0.68	83	72.1	30.4

### Descripción de las variables

$X_1$  será *POP\_CHING* que es el cambio poblacional de 1960 a 1970.

$X_2$  será *N\_EMPLD* que es el número de personas empleadas.

$Y$  será *PT\_POOR* que es el porcentaje de familias que viven por debajo del nivel de pobreza.

$X_3$  será *TAX\_RATE* que es la tasa de impuesto a la propiedad residencial y granjas.

$X_4$  será *PT\_PHONE* que es el porcentaje de viviendas con teléfonos.

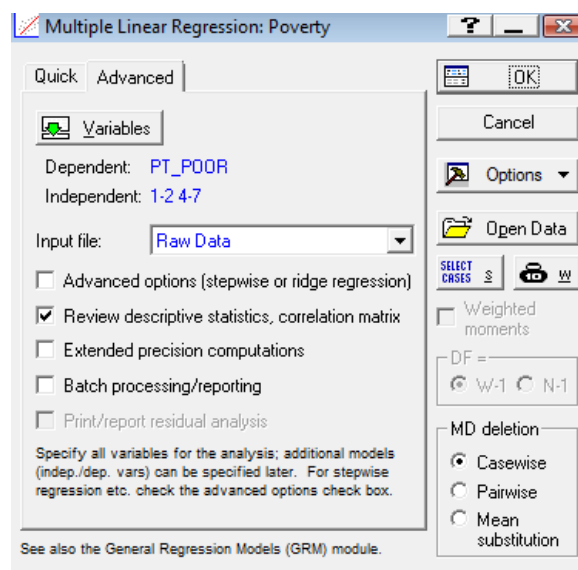
$X_5$  será *PT\_RURAL* que es el porcentaje de población rural.

$X_6$  será *AGE* que es la edad media.

Ahora bien, se analizará la correlación de la pobreza, es decir, las variables que mejor predicen el porcentaje de familias que viven por debajo del umbral de pobreza en un condado. Así que se tratará a la variable 3 (*Pt\_Poor*) como dependiente y el resto como independientes.

**Iniciando el análisis.** Hay que seleccionar *Multiple Regression* del menú *Statistics*. Especificar la ecuación de regresión haciendo clic en el botón *Variables* que está en el cuadro de diálogo de *Multiple Linear Regression*, se elige la pestaña *Quick* para mostrar el diálogo de selección de variables. Se escoge *Pt\_Poor* como la dependiente y todas las otras del archivo de datos se eligen como la lista de variables independientes y a continuación se hace clic en el botón OK. Además, en el cuadro de dialogo *Multiple Linear Regression* se selecciona la pestaña *Advanced* y luego se escoge el cuadrado denominado *Review descriptive statistics, correlation matrix*.

Tabla 3.1.2



Es bueno saber que este diálogo también está disponible a partir de básicamente todos los cuadros de diálogo de regresión múltiple, por lo que siempre puede volver a ver la estadística descriptiva para variables específicas.

**Distribución de las variables.** Primero hay que examinar la distribución de la variable dependiente *Pt\_Poor* a través de los condados. Hay que hacer clic en el botón de *Means & standard deviations* para mostrar la hoja de cálculo.

Tabla 3.1.3

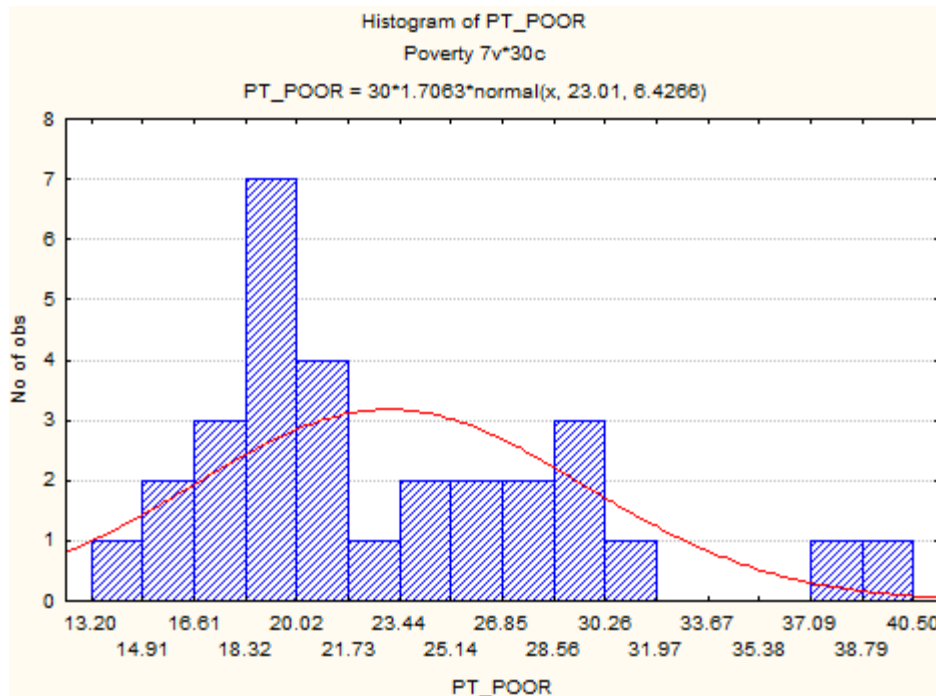
Variable	Means and Standard Deviations (Poverty)		
	Means	Std.Dev.	N
POP_CHNG	7.867	10.332	30
N_EMPLD	1548.667	2038.386	30
TAX_RATE	0.719	0.203	30
PT_PHONE	74.833	10.007	30
PT_RURAL	70.727	24.022	30
AGE	30.280	2.885	30
PT_POOR	23.010	6.427	30

Hay que seleccionar **Histograms** del menú de **Grafs** para producir el siguiente histograma de la variable *Pt\_Poor* (en el cuadro de diálogo resultante **2D Histograms** en la pestaña **Advanced**, luego se elige la opción *Categories* y se escribe 16 en el correspondiente campo de edición en el cuadro *Intervals* y se hace clic en el botón *OK*).

Como se puede ver a continuación, la distribución de esta variable se desvía ligeramente de la distribución normal. Los coeficientes de correlación pueden ser más grandes o más chicos considerablemente en caso de que existieran “outliers” extremos en los datos. Sin embargo, a pesar de que dos ciudades (las dos columnas de más a la derecha) tienen un mayor porcentaje de familias que viven por debajo del umbral de la pobreza de lo que cabría esperar en función de la normal, aún parecen estar lo suficientemente dentro del rango, pues la mayoría de las observaciones se encuentran adentro de esta curva.



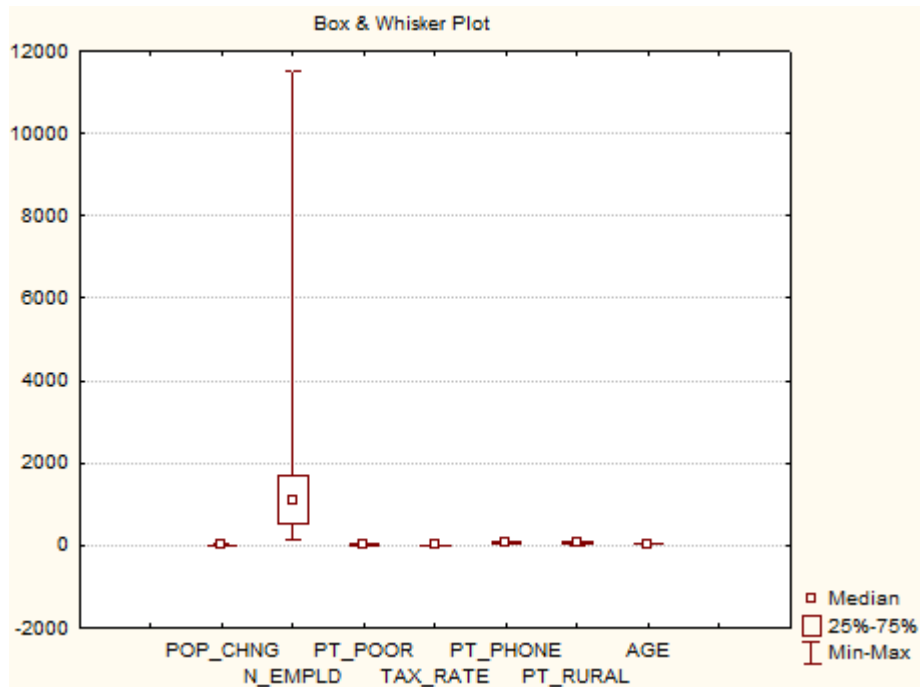
Gráfica 3.1.1



Esta decisión es subjetiva, se tiene que tener cuidado de que las observaciones no queden fuera de la media  $\pm 3$  veces la desviación estándar. En ese caso, es aconsejable repetir el análisis con y sin los “outliers” para garantizar que no afectan gravemente a la estructura de las variables explicativas. También se puede observar la distribución de esta variable haciendo clic en el botón *Box & whisker plot* sobre la pestaña *Advanced* que se encuentra en el cuadro de diálogo *Review Descriptive Statistics* y seleccionar todas las variables. A continuación, se escoge la opción *Median/Quartile/Range* sobre la ventana *Box-Whisker Type*, y luego, se da clic en el botón *OK* para producir la *Box and Whisker Plot*, que se muestra a continuación.

Un diagrama de caja es un gráfico, basado en cuartiles, mediante el cual se visualiza un conjunto de datos. Está compuesto por un rectángulo, la "caja", y dos brazos. Este suministra información sobre los valores mínimo y máximo, los cuartiles, sobre la existencia de posibles valores atípicos y la simetría de la distribución.

Gráfica 3.1.2



Se advierte que la covariable  $X_2$ , representada para el número de personas empleadas, tiene una mayor dispersión con valores muy diferentes a las otras variables independientes.

**Gráficas de dispersión.** Si se tiene una hipótesis *a priori* acerca de la relación entre variables específicas en este punto esto puede ser instructivo para representar los respectivos puntos de dispersión. Por ejemplo, analizar la relación entre los cambios de la población y el porcentaje de familias que viven por debajo del nivel de pobreza. Parece razonable pronosticar que la carencia conduzca a la emigración, por lo tanto, debería existir una correlación negativa entre el porcentaje por debajo del nivel de pobreza y los cambios de la población.

Regresando a la ventana de diálogo *Review Descriptive Statistics*, hay que hacer clic en el botón de *Correlations*, que se encuentra sobre la pestaña *Quick* para mostrar la hoja de cálculo con la matriz de correlación.

Tabla 3.1.4

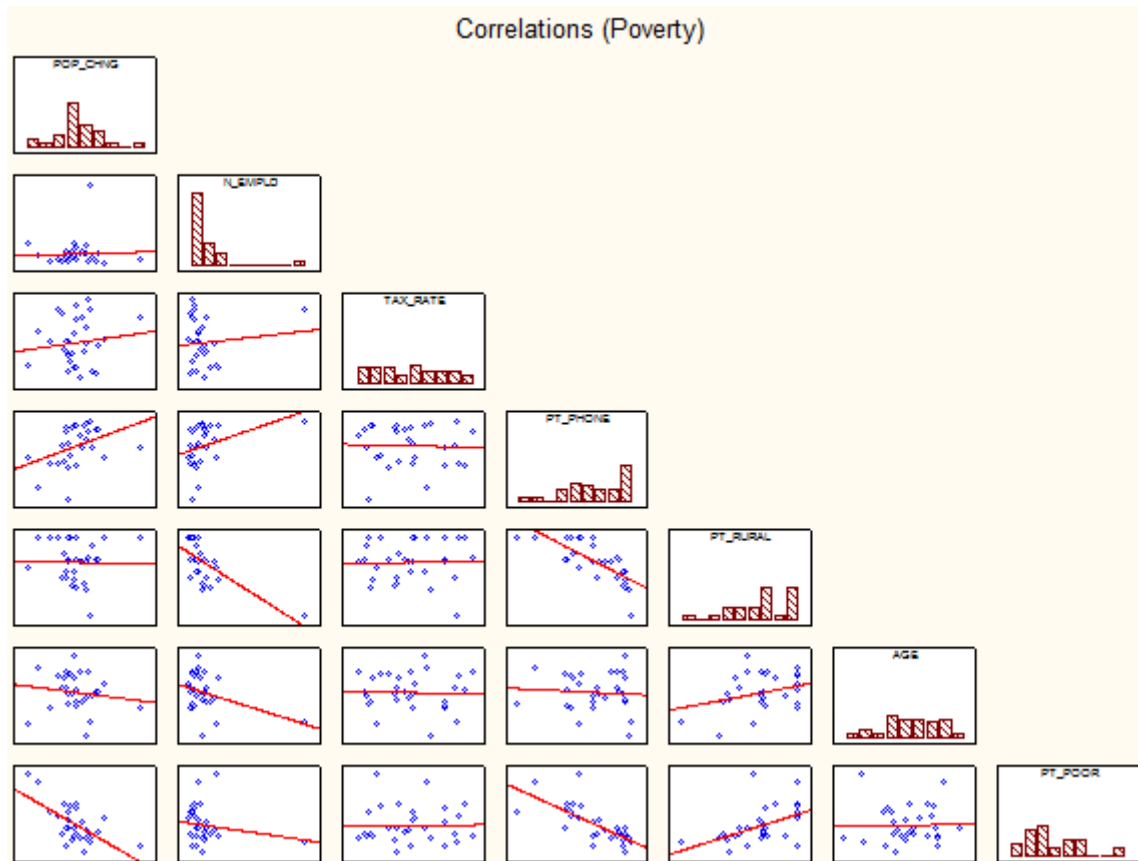
Variable	POP_CHNG	N_EMPLD	TAX_RATE	PT_PHONE	PT_RURAL	AGE	PT_POOR
POP_CHNG	1.000000	0.040296	0.130987	0.378333	-0.018770	-0.146903	-0.649111
N_EMPLD	0.040296	1.000000	0.103548	0.355492	-0.657735	-0.363706	-0.167687
TAX_RATE	0.130987	0.103548	1.000000	-0.038022	0.023300	-0.046692	0.008984
PT_PHONE	0.378333	0.355492	-0.038022	1.000000	-0.748583	-0.078356	-0.733467
PT_RURAL	-0.018770	-0.657735	0.023300	-0.748583	1.000000	0.314420	0.512603
AGE	-0.146903	-0.363706	-0.046692	-0.078356	0.314420	1.000000	0.020675
PT_POOR	-0.649111	-0.167687	0.008984	-0.733467	0.512603	0.020675	1.000000

En esta tabla se observa que la correlación entre las variables independientes y la dependiente son bastante diferentes. Por ejemplo la que se halla entre variable *PT\_POOR* y *AGE* no es nada buena, pues ésta es de 0.020675, una cantidad muy pequeña a diferencia de la existente con las variables *POP\_CHING* y *PT\_PHONE* las cuales tienen una correlación con *PT\_POOR* de -0.649111 y de -0.733467, respectivamente (más adelante se verá si esta última es realmente significativa con otra prueba).

También es posible observar en esta tabla las correlaciones entre las covariables, por ejemplo, la existente entre la variable *PT\_RURAL* y *PT\_PHONE* que es bastante alta (con un coeficiente de -0.748538), lo cual podría causar problemas si se utilizaran ambas variables como explicativas para el análisis. A esta última relación se le conoce como multicolinealidad.

Las correlaciones entre las variables también pueden ser mostradas en una matriz de dispersión de variables, se puede seleccionar y producir dando clic en el botón *Matrix plot of correlations*, que está sobre la **Review Descriptive Statistics** en la opción **Advanced** para luego seleccionar las variables deseadas.

Gráfica 3.1.3



Esta gráfica múltiple sirve para determinar si entre las variables explicativas existe algún tipo de colinealidad, además para observar cuál de las covariables será la que pueda aportar más al análisis de regresión, así mismo determina si la relación lineal (u otro tipo de dependencia) propuesta es la correcta en cada variable independiente.

En este caso se observa que la variable *PT\_POOR* tiene una correlación más o menos lineal con las variables *POP\_CHING* y *PT\_RURAL*.

**Especificación de la Regresión Múltiple.** Ahora se hace clic en el botón *OK* en la ventana de diálogo *Review Descriptive Statistics* para llevar a cabo el análisis de regresión, se mostrará la ventana de diálogo *Multiple Regression Results* para obtener los resultados del análisis. Una regresión estándar (que incluye la intercepción) se abrirá.



La tabla anterior muestra la normalización de los coeficientes de regresión (Beta) y los coeficientes no estandarizados de regresión (B). La magnitud de estos coeficientes beta permiten comparar la contribución relativa de cada variable explicativa en la predicción de la respuesta. Como se ha señalado anteriormente es evidente que en la tabla, las covariables *Pop\_Chng*, *Pt\_Rural*, y *N\_Empld* son los predictores más importantes de la pobreza, de los cuales sólo las dos primeras de estas son consideradas como estadísticamente significativas. El coeficiente de regresión para *Pop\_Chng* es negativo, es decir, que el menor aumento de la población causará que sea mayor el número de familias que viven por debajo del umbral de la pobreza en los respectivos condados. Además la cifra para *Pt\_Rural* es positiva, es decir, que mientras mayor sea el porcentaje de la población rural, mayor es el nivel de pobreza.

La ANOVA es:

Tabla 3.1.7

Effect	Sums of Squares	df	Mean Squares	F	p-level
Regress.	932.065	6	155.3441	13.44909	0.000002
Residual	265.662	23	11.5505		
Total	1197.727				

La tabla anterior especifica el valor de cada una de las sumas; debida a la regresión ( $SS_R = \hat{\beta}_1 S_{xy}$ ), correspondiente a los residuales ( $SS_E = S_{yy} - \hat{\beta}_1 S_{xy}$ ) y al total ( $S_{yy}$ ). También se pueden observar los cuadrados medios ( $MS_R$  y  $MS_E$ ), la prueba *F* (por *Fisher*) que es la división de los cuadrados medios de la regresión entre los cuadrados medios de los residuos y por último el nivel de significancia (también llamado p-level o p-value). El estadístico *F* a lo que se refiere es que aproximadamente los efectos de la regresión son 13.45 veces mayores que las consecuencias debidas a la fluctuación aleatoria (o a los errores).

En esta tesis sólo se trabajará con problemas que involucren una regresión lineal simple, es decir, sólo se considerará una variable explicativa. Por lo que se ocupará la que tuvo

un p-level más pequeño en la prueba de significancia, en este caso sería la variable *POP\_CHING* que resultó con un p-level igual a 0.000064. Esta decisión es fundada a que si por ejemplo, se especifica que se quiere cometer a lo más un error del 1% para la prueba de significancia del coeficiente de regresión de esta covariable, entonces se tendría que el p-value es menor que dicho error y por lo tanto se rechazaría la hipótesis de nulidad de que este coeficiente sea igual a cero (por lo que existe asociación entre las variables independiente y *POP\_CHING*). Los otros coeficientes significativos bajo esta condición son: la intersección y *PT\_RURAL*.

**Modelo reducido con una variable.** A continuación se harán los mismo procesos que se hicieron anteriormente, para analizar la regresión lineal simple, posteriormente se verá si existen “outliers” (puntos atípicos) y además las pruebas de residuales entre otras más.

Tabla 3.1.8

Regression Summary for Dependent Variable: PT_POOR (Poverty)						
R= .64911115 R <sup>2</sup> = .42134529 Adjusted R <sup>2</sup> = .40067905						
F(1,28)=20.388 p<.00010 Std. Error of estimate: 4.9752						
N=30	Beta	Std.Err. of Beta	B	Std.Err. of B	t(28)	p-level
Intercept			26.18623	1.148871	22.79301	0.000000
POP_CHNG	-0.649111	0.143758	-0.40374	0.089416	-4.51532	0.000104

Cuando se realiza la regresión con solamente una variable significativa se tiene un coeficiente de determinación del 42.13%, es importante notar que disminuye dicho coeficiente al original del modelo de 6 variables, pero esto no afectará los fines prácticos de este trabajo. Como ya se sabía la variable *POP\_CHING* es significativa y esto precisamente se demuestra en la tabla anterior, aunque se tenga que el p-level aumentó a 0.000104 (sigue siendo un valor significativo para la prueba).

**Intervalos de confianza para la intersección y la pendiente.** Con una confianza del 95% el valor real de la intersección y de la pendiente están dentro de los intervalos (23.81, 28.62) y (-0.59, -0.21), respectivamente. Dado que en ninguno de los dos intervalos está contenido el cero, no se realizarán pruebas de hipótesis para ver si son

cero o no lo son, es decir, en esta regresión los dos coeficientes son significativos, por lo que estas dos variables se pueden relacionar mediante una ecuación lineal.

Lo siguiente a realizar es obtener la tabla ANOVA, en la cual se observará todo lo referente a la suma de los cuadrados debidos a la regresión, a los residuales y al total, así como los cuadrados medios, la prueba de  $F$  y el p-value.

Tabla 3.1.9

Effect	Sums of Squares	df	Mean Squares	F	p-level
Regress.	504.657	1	504.6566	20.38810	0.000104
Residual	693.070	28	24.7525		
Total	1197.727				

Se puede apreciar que los cuadrados medios asociados a la regresión son mucho mayores que los de los errores, por lo que la prueba  $F^{11}$  resultó muy buena, dado que asegura que la fuente de variación debida a la regresión es aproximadamente 20.39 veces mayor que la de los efectos debidos a los residuos.

La ecuación de regresión hasta este momento quedaría de la siguiente forma:

$$Y = 26.18623 - 0.40374X_1.$$

La tabla y la matriz de correlaciones dice lo siguiente:

Tabla 3.1.10

Variable	POP_CHNG	PT_POOR
POP_CHNG	1.000000	-0.649111
PT_POOR	-0.649111	1.000000

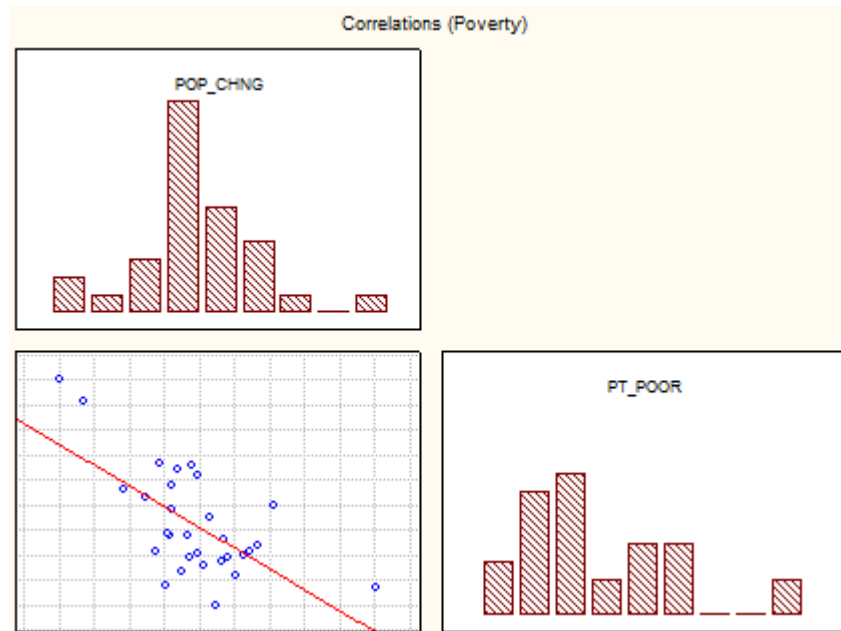
<sup>11</sup> La prueba de  $F$  y la de  $t$  de Student están relacionadas, ya que la segunda de estas elevada al cuadrado es igual a la primera.



Se observa que existe una correlación de  $-0.649111$  entre la variable dependiente y la independiente. Se trata de una asociación bastante buena, aunque se pudiera mejorar si se hallaran “outliers” y se eliminaran del análisis.

Una forma más visual de observar esta correlación es haciendo la gráfica de dispersión entre ambas variables.

Gráfica 3.1.4



**Análisis de los residuales.** Después de ajustar una ecuación de regresión, siempre se debe examinar los resultados de predicción y los residuos. Por ejemplo, los “outliers” que pueden suponer resultados equivocados y llevar a conclusiones erróneas. Desde la ventana de diálogo *Multiple Regression Results* y en la pestaña *Residuals/assumptions /prediction*, se hace clic en el botón *Perform residual analysis* para continuar con la ventana de diálogo *Residual Analysis*, para después analizar las gráficas de residuos.

**Gráficas de los residuos.** En general, se debería al menos examinar la estructura de los residuos normalizados para identificar cualquier punto atípico. Hay que ir a la pestaña *Residuals* y hacer clic en el botón *Casewise plot of residuals*, por default, el *Raw residuals* (residuales no estandarizados) será dibujado en una tabla, sin embargo, también se pueden seleccionar otros residuales estadísticos en el recuadro *Type of residual* que tiene varias opciones de residuales.

Tabla 3.1.11

Raw Residual (Poverty)									
Dependent variable: PT_POOR									
Case name	Raw Residuals				Observed Value	Predicted Value	Residual	Standard Pred. v.	
	-3s	.	0	.	+3s				
Benton	.	.	*	.	.	19.00000	20.65498	-1.65498	-0.56454
Cannon	.	.	*	.	.	26.20000	26.50923	-0.30923	0.83883
Carrol	.	.	*	.	.	18.10000	22.31032	-4.21032	-0.16773
Cheatheam	.	.	.	*	.	15.40000	10.03658	5.36342	-3.10997
Cumberland	.	.	.	*	.	29.00000	22.79077	6.20923	-0.05255
DeKalb	.	.	*	.	.	21.60000	24.77314	-3.17314	0.42266
Dyer	.	.	*	.	.	21.90000	24.97501	-3.07501	0.47105
Gibson	.	.	*	.	.	18.90000	23.31967	-4.41967	0.07423
Greene	.	.	*	.	.	21.10000	20.93760	0.16240	-0.49679

La escala utilizada en la gráfica de la columna del extremo izquierdo se encuentra en términos de sigma, es decir, la desviación estándar de los residuos. Si uno o varios casos quedan fuera de los límites de sigma de  $\pm 3$ , probablemente se deberían excluir los respectivos casos (lo cual se hace a través de las condiciones de selección de casos) y habrá que ejecutar el análisis otra vez para asegurarse de que los principales resultados no fueron sesgados por estos “outliers”.

**Gráfica de outliers.** Una manera rápida de identificar los “outliers” es utilizar el botón *Casewise plot of outliers* que está en la pestaña **Outliers**. Se puede también dibujar todos los residuos estandarizados que quedan fuera de los límites de sigma  $\pm 2$  o dibujar los 100 casos más extremos, como se especifica en el recuadro *Type of outlier* que está en la pestaña **Outliers**. Cuando se selecciona el *Standard residual* ( $> 2 * \text{sigma}$ ) del botón de opción, no hay puntos extremos detectados en el ejemplo actual.

**Distancias de Mahalanobis.** La mayoría de los libros de texto de estadística dedican una discusión a la cuestión de los valores atípicos y los residuos en relación con la variable dependiente. Sin embargo, el papel de los valores extremos en la variable independiente es a menudo pasado por alto. En esta cuestión se tiene una lista de variables que participan con diferentes pesos (los coeficientes de regresión) en la predicción de la variable dependiente. Se puede pensar en las covariables como la definición de un espacio multidimensional en el que cada observación puede ser localizada. Por ejemplo, si se tienen dos variables independientes con los mismos coeficientes de regresión, entonces se lograría construir un diagrama de dispersión de ambas, y se podría observar cada punto de cada variable en un dibujo por separado.

Entonces se dibujaría un punto para la media de ambas variables y se calcularían las distancias de cada observación de esta media (ahora llamado centroide) en el espacio bidimensional, lo que es la idea conceptual detrás de calcular las distancias de Mahalanobis. Ahora, hay que observar éstas (ordenados por tamaño) para identificar los casos extremos en el lado de variable independiente. Hay que seleccionar el botón de opción *Mahalanobis distances* en el cuadro de *Type of outlier*, a continuación, se hace clic en el botón *Casewise plot of outliers*. El resultado se mostrará en la tabla de distancias de Mahalanobis mandados en orden descendente.

Tabla 3.1.12

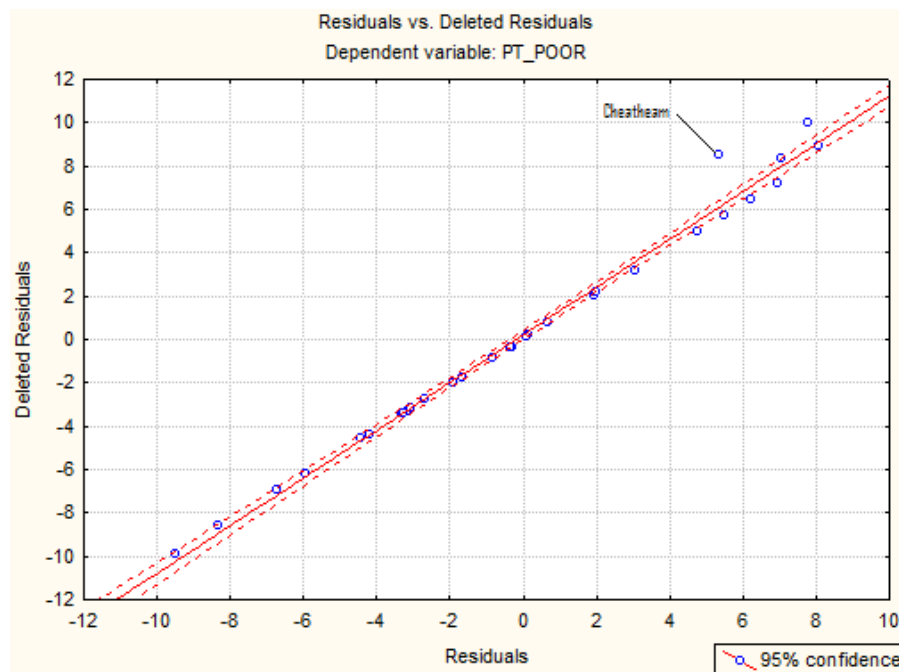
Mahalanobis distances: PT_POOR (Poverty)						
Sorted						
Case name	.001	.	.	.	.	9.67
	Observed Value	Predicted Value	Residual	Standard Pred. v.		
Cheatheam	15.40000	10.03658	5.36342	-3.10997	*	
Haywood	40.50000	32.72684	7.77316	2.32930		*
Jackson	38.00000	30.95038	7.04962	1.90345	*	
Houston	25.40000	17.34430	8.05570	-1.35817	*	
Morgan	27.30000	28.12419	-0.82419	1.22597	*	
Weakley	20.50000	18.51515	1.98485	-1.07750	*	
Humphreys	19.70000	18.99964	0.70037	-0.96136	*	
Cannon	26.20000	26.50923	-0.30923	0.83883	*	
Sevier	19.20000	19.52450	-0.32450	-0.83554	*	
Shelby	16.80000	20.04936	-3.24937	-0.70972	*	


Hay que tener en cuenta que el condado de *Cheatheam* (en la primera línea) parece un tanto extremo, en comparación con las demás ciudades del gráfico. Si se aprecian los datos de los valores observados se observará que, de hecho, el condado *Cheatheam* es el territorio que sufrió un mayor cambio poblacional de 1960 a 1970 (la variable *POP\_CHING*). Probablemente, hubiera sido conveniente expresar esos números en porcentajes y no en absolutos, y en ese caso, la distancia de Mahalanobis de la región de *Cheatheam* posiblemente no habría sido tan grande del resto de los condados en la muestra. Pero se tendrán que hacer otras pruebas para saber si verdaderamente este lugar es un valor atípico.

**Residuales borrados.** Otra opción muy importante que permite una evaluación de la gravedad del problema de los “outliers” es el *Deleted residuals*. Ésta es la normalización de los residuales para el caso respectivo que se pudiera obtener si el “outlier” se excluyera del análisis. Hay que recordar que el procedimiento de regresión lineal simple ajusta una línea recta para expresar la relación entre la variable

independiente y la dependiente. Si un caso es un valor atípico entonces puede haber una tendencia de la línea de regresión a ser "jalada" por este "outlier". Como resultado de ello, si los respectivos casos fueran excluidos, surgiría un nuevo coeficiente de regresión y una línea totalmente diferente. Por lo tanto, si los residuales borrados son muy diferentes de la normalización de los residuales, se tienen razones suficientes para creer que el análisis de regresión es muy sesgado por los respectivos casos. En este ejemplo, el residual borrado del condado de *Cheatheam* no es un valor atípico (pero se realizará una prueba más para corroborarlo). Se puede dibujar los residuales contra los residuales borrados haciendo clic en el botón *Residuals vs. deleted residuals* sobre la pestaña de **Scatterplots** lo que producirá un diagrama de dispersión de estos valores. La dispersión mostrará los valores atípicos (si los hay).

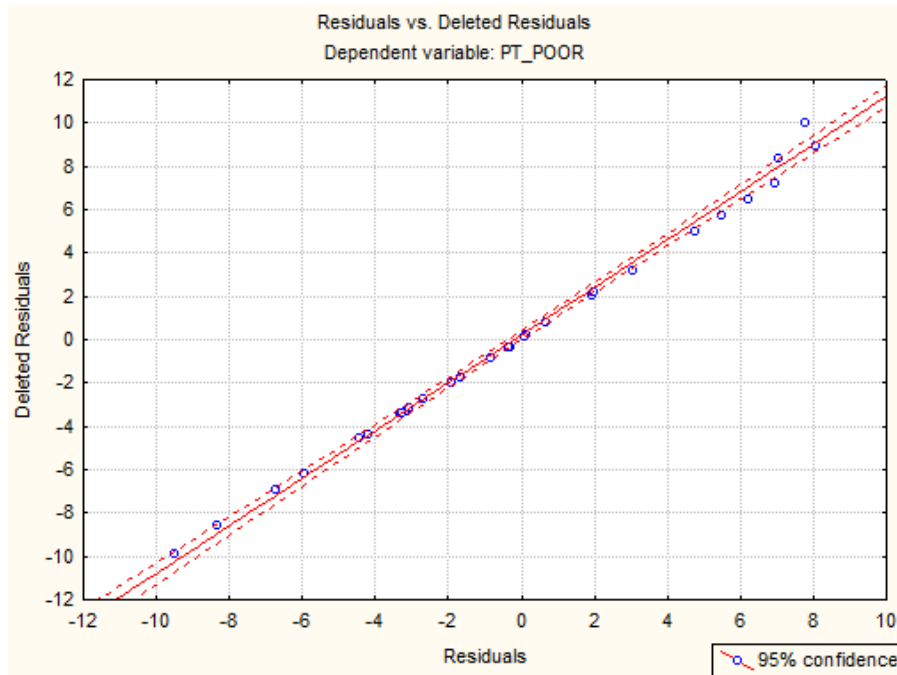
Gráfica 3.1.5



El programa STATISTICA ofrece una herramienta de eliminación de "outliers" interactiva llamada *Brushing Tool*  con el fin de experimentar con la eliminación de los valores atípicos; al instante se verá su influencia en la línea de regresión. Cuando la herramienta está activada, el cursor cambia a una cruz, y el dialogo **Brushing** se mostrará junto al gráfico. Interactivamente se pueden eliminar los puntos mediante la selección de la casilla de verificación *Auto Update* y el botón *Turn OFF* de opción en

el cuadro *Action*, y a continuación, se hace clic en el punto que se va a quitar con la cruz. Entonces se podrá quitar automáticamente de la gráfica ese punto.

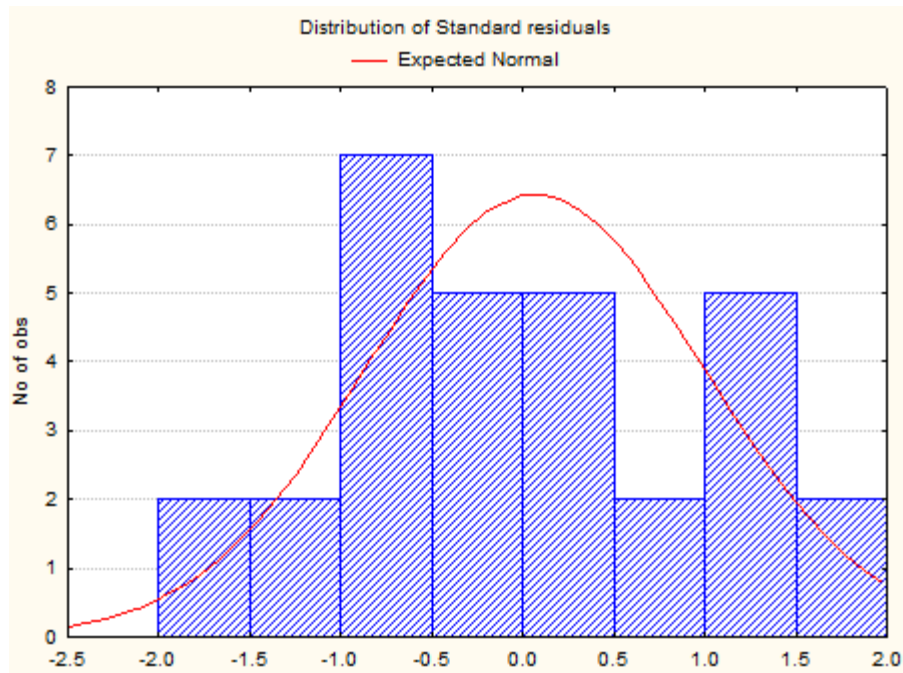
Gráfica 3.1.6



Se puede regresar el punto eliminado haciendo clic en el botón *Reset All* en el diálogo de *Brushing*. Como puede verse en la gráfica anterior, ya no está el condado de *Cheatham*, pero además se observa que precisamente no es un punto influyente, pues la recta no es “jalada”.

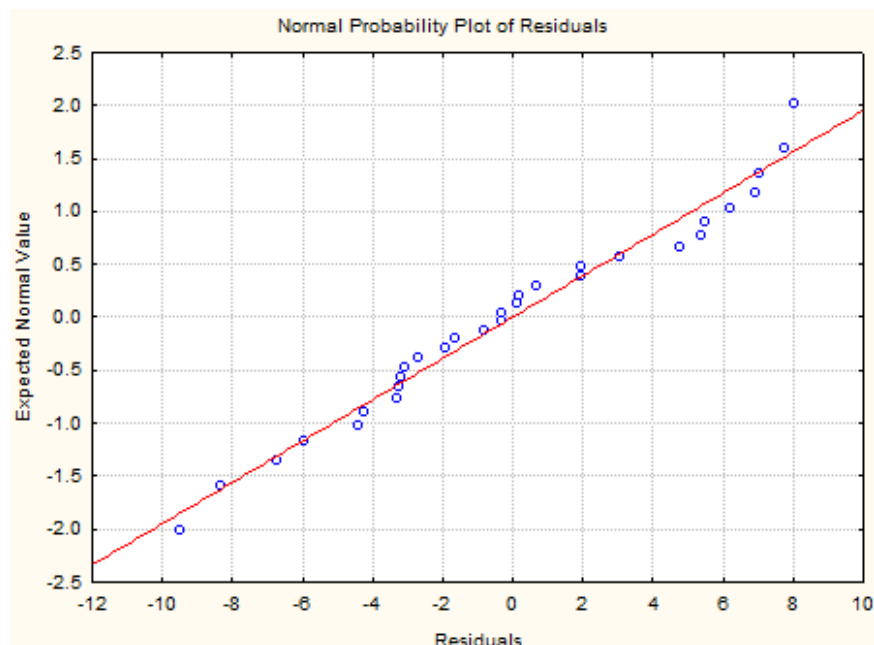
**Distribución de los residuales normalizados.** En la ventana de diálogo *Residual Analysis* y en la pestaña de *Residuals* se encuentra la opción de *Histogram of residuals*, si se selecciona del cuadro de *Type of residual* la opción de *Standard residual* se obtiene el siguiente gráfico. Se podrá ver que no se alcanza plenamente la normalidad de estos residuos, pero la mayoría de las columnas se agrupan dentro de la curva de la normal.

Gráfica 3.1.7



**Gráfica de Normalidad.** Hay muchos gráficos adicionales disponibles en el cuadro de diálogo *Residual Analysis*. Como se señaló anteriormente, la regresión lineal simple supone en los errores una distribución normal. Si estas hipótesis no se cumplen, su conclusión final puede no ser exacta. La gráfica de normalidad de los residuos dará una sospecha de si es grave o no la violación del supuesto. Hay que hacer clic en el botón *Normal plot of residuals* (en la pestaña *Probability plots*) para mostrar este gráfico.

Gráfica 3.1.8



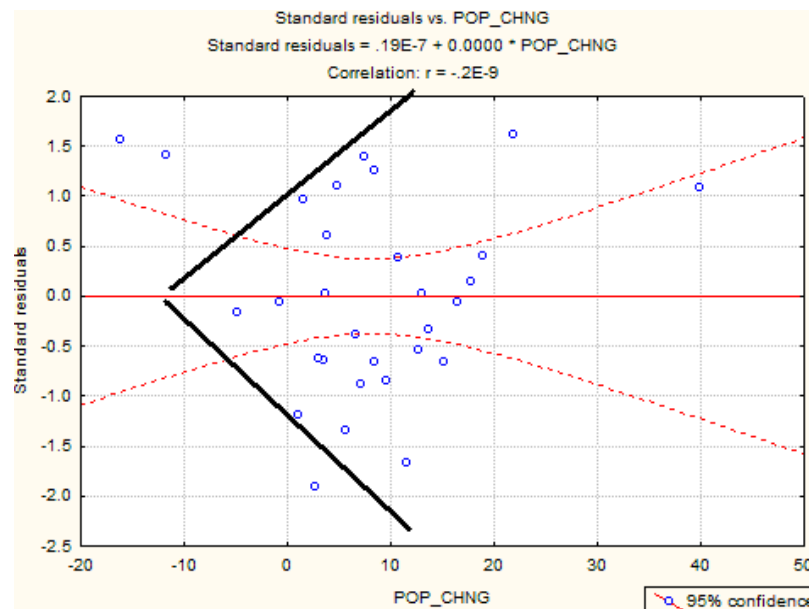
La gráfica anterior se construye de la siguiente manera. En primer lugar los residuos se ordenaron por rango y a partir de estos, los valores  $z$  pueden ser calculados (es decir, los valores estándar de la distribución gaussiana), basados en el supuesto de que los datos provienen de una distribución normal<sup>12</sup>. Estos valores  $z$  se representan gráficamente sobre el eje  $y$  en la gráfica.

Si se observan los residuales (trazados en el eje  $x$ ) todos los valores siguen de cerca la línea recta, con esto se podría creer que se distribuyen normalmente. Si los residuos no se distribuyen normalmente, entonces se desvían de la esta. Igualmente, los valores atípicos pueden hacerse evidentes en la gráfica.

Si hay una falta general de ajuste y los datos parecen formar un patrón claro (por ejemplo, una forma de S) en torno a la línea, entonces, la variable dependiente pudiera ser transformada de algún modo (una transformación de logaritmo podría "jalar" la cola de la distribución). Sin embargo, con demasiada frecuencia, los investigadores simplemente aceptan sus datos en valor nominal sin la comprobación de la idoneidad de sus supuestos, lo que lleva a conclusiones muchas veces, erróneas.

### Varianza Constante.

Gráfica 3.1.9

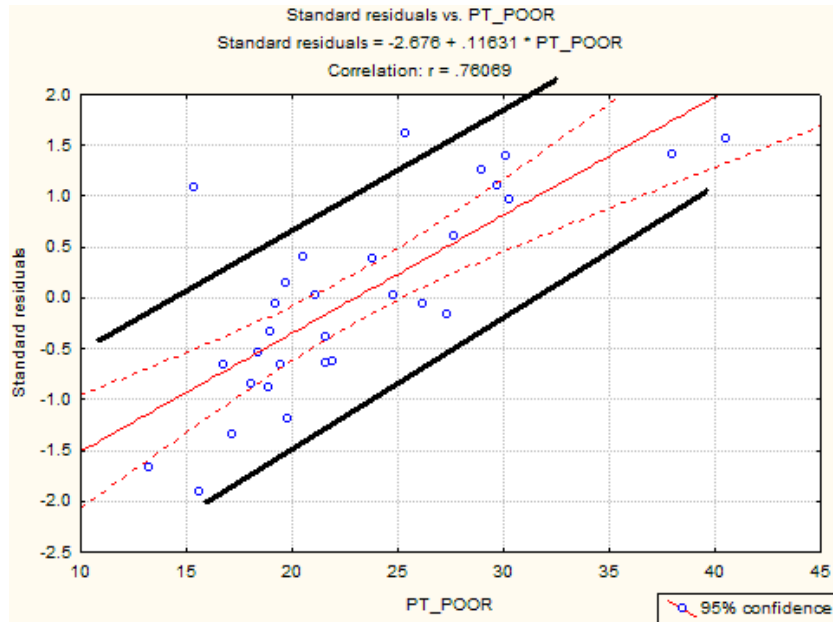


<sup>12</sup> Entre las ecuaciones para calcular los valores  $z$ , se encuentra  $(i - 0.375)/(n + .25)$ , donde  $i$  es el contador de los residuos ordenados y  $n$  es el número de observaciones.

En la gráfica anterior existe un efecto de embudo (es decir, no existe una varianza constante). Si se utilizó el método de mínimos cuadrados para la obtención de los estimadores, se podría mejorar esto si se obtuvieran por mínimos cuadrados ponderados.

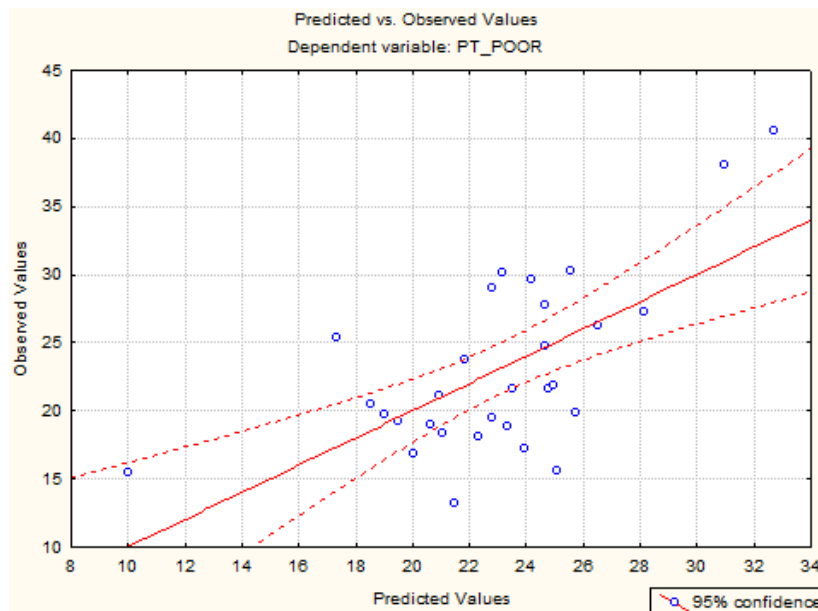
La gráfica de los residuos estandarizados de la variable dependiente.

Gráfica 3.1.10



En el gráfico siguiente se nota que los puntos se alejan un poco de recta, por lo que se pudiera pensar que los valores de  $y$  no son predichos por  $\hat{y}$ .

Gráfica 3.1.11





### 3.2 Ejemplo 2

Este modelo se basa en el archivo de datos de ejemplos del paquete STATISTICA, y se llama *Job\_prof.sta* (de Neter, Wasserman y Kutner, 1989, página 473). Este archivo de datos se abre a través de *File-Open* del menú de ejemplos, que se encuentra en la carpeta de datos. Las cuatro primeras variables (*TEST1-TEST4*) representan diferentes pruebas de aptitud que se administraron a cada uno de los 25 solicitantes de nivel de entrada para los puestos de oficina en una empresa. Independientemente de sus resultados en las pruebas, todos los candidatos fueron contratados. Una vez que su período de prueba había expirado, cada empleado fue evaluado y se le asignó un empleo de acuerdo a la calificación obtenida en los exámenes (variable *Job\_prof*).

Los datos se muestran en la siguiente tabla.

Tabla 3.2.1

Job proficiency data set from Neter, Wasserman, & Kutner, 1989						
	1	2	3	4	5	
	TEST1	TEST2	TEST3	TEST4	JOB_PROF	
1	86	110	100	87	88	
2	62	97	99	100	80	
3	110	107	103	103	96	
4	101	117	93	95	76	
5	100	101	95	88	80	
6	78	85	95	84	73	
7	120	77	80	74	58	
8	105	122	116	102	116	
9	112	119	106	105	104	
10	120	89	105	97	99	
11	87	81	90	88	64	
12	133	120	113	108	126	
13	140	121	96	89	94	
14	84	113	98	78	71	
15	106	102	109	109	111	
16	109	129	102	108	109	
17	104	83	100	102	100	
18	150	118	107	110	127	
19	98	125	108	95	99	
20	120	94	95	90	82	
21	74	121	91	85	67	
22	96	114	114	103	109	
23	104	73	93	80	78	
24	94	121	115	104	115	
25	91	129	97	83	83	

## Descripción de las variables

$X_1$  será *TEST1* que son los resultados de la primera prueba de aptitud.

$X_2$  será *TEST2* que son los resultados de la segunda prueba de aptitud.

$X_3$  será *TEST3* que son los resultados de la tercera prueba de aptitud.

$X_4$  será *TEST4* que son los resultados de la cuarta prueba de aptitud.

Y será *JOB\_PROF* que es el dominio del trabajo después del periodo de prueba.

Entonces se procederá a revisar cual de las covariables (*TEST1-TEST4*) predice mejor el dominio del trabajo, por lo tanto, la variable dependiente será *JOB\_PROF*.

**Iniciando el análisis.** Se hace exactamente lo mismo que se hizo en el primer ejemplo.

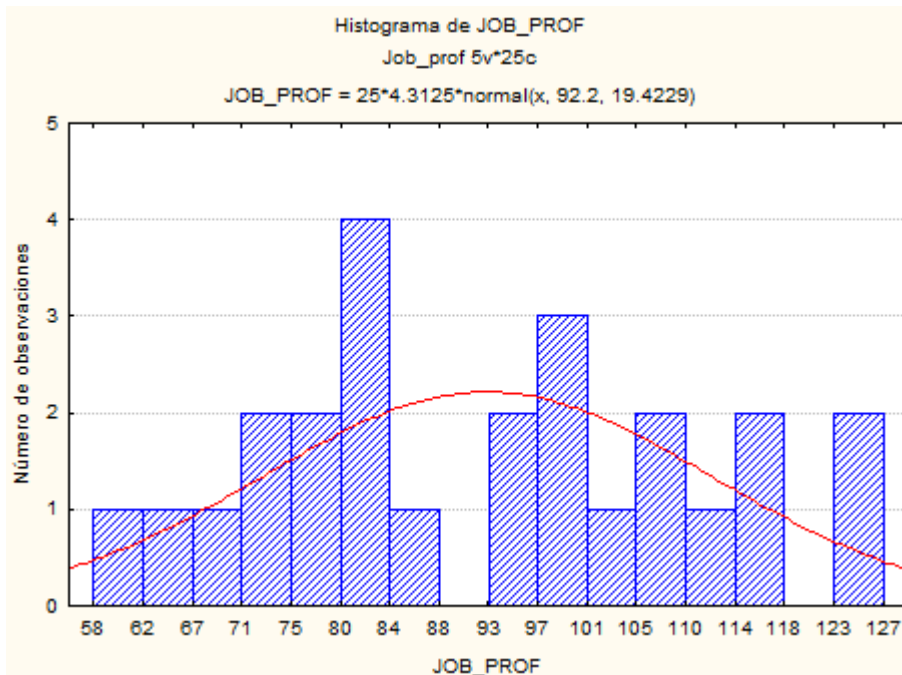
**Distribución de las variables.** Como ya se sabe, primero hay que examinar la distribución de la variable dependiente *JOB\_PROF* a través de las cuatro pruebas. Aquí se observan las medias y las desviaciones estándares de las distintas variables.

Tabla 3.2.2

Variable	Medias y Desv. Est. (Job_prof)		
	Medias	Desv. Est.	N
TEST1	103.3600	20.29754	25
TEST2	106.7200	17.29287	25
TEST3	100.8000	8.85532	25
TEST4	94.6800	10.67599	25
JOB PROF	92.2000	19.42292	25

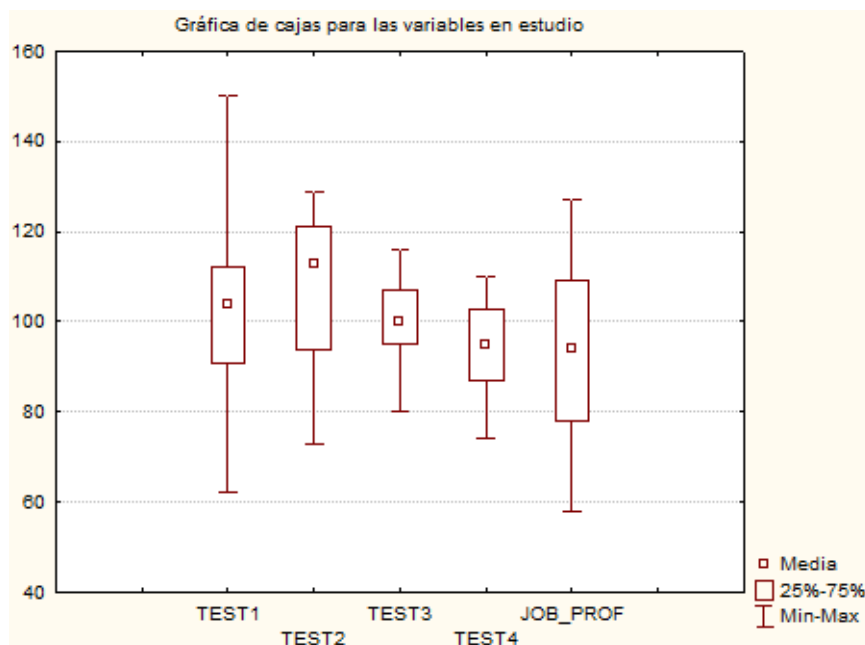
Por ejemplo, la media y la desviación estándar de la variable independiente es de 92.2 y de 19.42292, respectivamente. Además se aprecia el número de observaciones, que en este caso es de 25 para todas las variables. La mayor desviación estándar de las variables la tiene *TEST1*, ya que ésta es de 20.29754; y la más pequeña de aquellas es la de *TEST3* pues resultado de 8.85532.

Gráfica 3.2.1



El histograma muestra que la distribución de la variable *JOB\_PROF* se desvía de la normal. De hecho, en el centro en donde se encuentra la media, hay un intervalo en donde no hay datos. Sin embargo, la mayor parte de ellos se encuentran dentro del rango, aunque exista dispersión, pues es claro que los datos se alejan de la media. Hay más empleados que obtuvieron un menor puntaje y menos personas que consiguieron un mayor puntaje en las pruebas.

Gráfica 3.2.2



Se observa que todas las variables tienen bastante dispersión, con valores muy diferentes de una a otra. Las medias son muy distintas entre las variables, al igual que los máximos y mínimos.

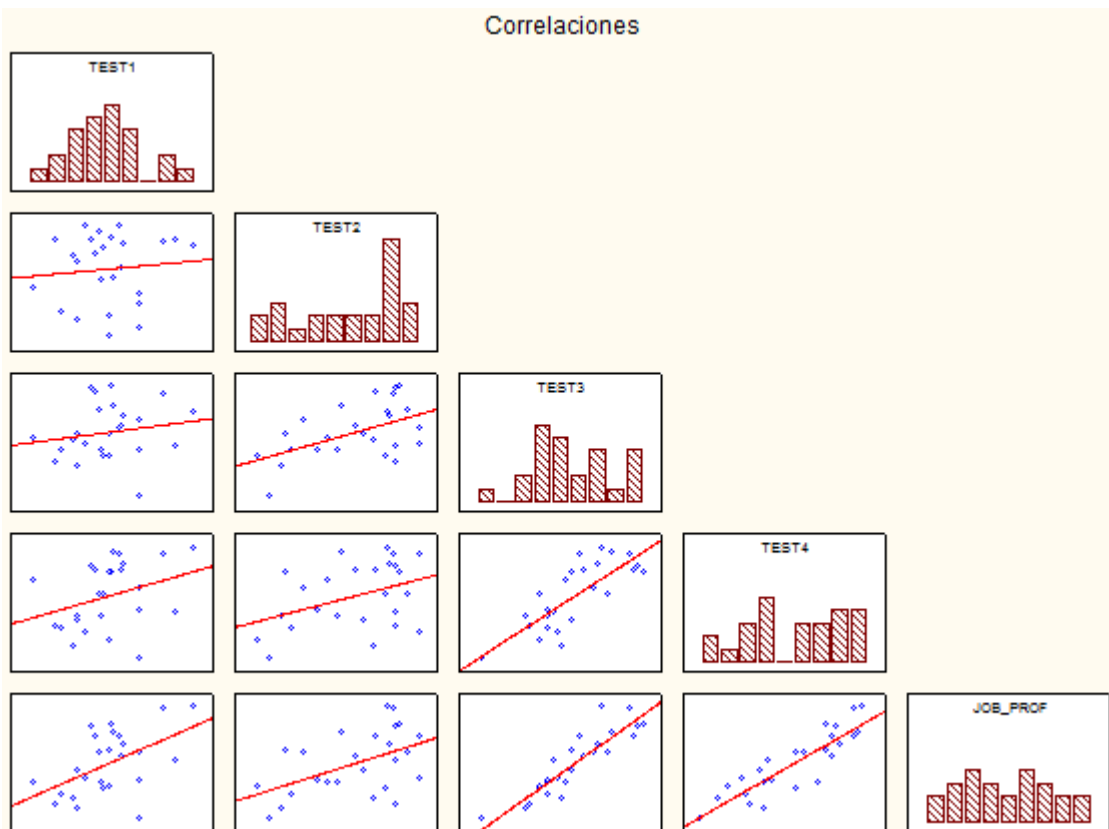
Ahora se verán las correlaciones entre todas las variables, con la siguiente tabla y con el gráfico de matrices de correlación.

Tabla 3.2.3

Variable	Correlaciones (Job_prof)				
	TEST1	TEST2	TEST3	TEST4	JOB_PROF
TEST1	1.000000	0.102269	0.180769	0.326663	0.514411
TEST2	0.102269	1.000000	0.519045	0.396710	0.497006
TEST3	0.180769	0.519045	1.000000	0.782038	0.897064
TEST4	0.326663	0.396710	0.782038	1.000000	0.869387
JOB_PROF	0.514411	0.497006	0.897064	0.869387	1.000000

Las correlaciones entre las variables varían bastante, la mejor de estas se observa entre *TEST3* y *JOB\_PROF*, pues resultó ser de 0.869387.

Gráfica 3.2.3



Como se puede observar en la gráfica anterior, la correlación entre la variable de respuesta y las covariables son distintas. Las dos independientes que más correlación presentan con la dependiente son *TEST3* y *TEST4*, aunque también las otras presentan una buena relación, pero no tan grande como las dos primeras señaladas. A pesar de esto, la colinealidad entre estas dos variables explicativas es bastante grande, por lo que si se quisiera hacer un análisis de regresión múltiple no sería recomendable utilizar ambas.

### Revisión de los resultados

Tabla 3.2.4

Resumen de la regresión para la variable dependiente: JOB_PROF						
R= .98127051 R²= .96289181 Ajustada R²= .95547018						
F(4,20)=129.74 p<.00000 Desviación Estándar de la estimación: 4.0986						
N=25	Beta	Desv. Est. de Beta	B	Desv. Est. de B	t(20)	p-level
<b>Intersección</b>			-124.382	9.941063	-12.5119	0.000000
TEST1	0.309042	0.045951	0.296	0.043971	6.7254	0.000002
TEST2	0.042992	0.050408	0.048	0.056617	0.8529	0.403826
TEST3	0.595438	0.074813	1.306	0.164091	7.9591	0.000000
TEST4	0.285723	0.072524	0.520	0.131943	3.9397	0.000810

Las pruebas de significancia dicen que las variables independientes *TEST1*, *TEST3* y *TEST4* son significativas. Pero se trabajará con la que presentó un mejor coeficiente (p-level menor) y mayor correlación, es decir, con la variable *TEST 3*.

La ANOVA de esta regresión es la siguiente.

Tabla 3.2.5

Análisis de Varianza; Var. Dep: JOB_PROF					
Fuente de variación	Suma de Cuadrados	Grados de Libertad	Cuadrados Medios	F	p-level
<b>Regresión</b>	<b>8718.022</b>	4	2179.506	129.7411	0.000000
Residuales	335.978	20	16.799		
Total	9054.000				

La prueba *F* asegura que la fuente de variación debida a la regresión es 129.7411 veces mayor que la de los residuos, por lo que en este aspecto se trata de un buen modelo.

**Modelo reducido con una variable.** Los pasos a seguir son los mismos que en el ejemplo anterior.

Tabla 3.2.6

Resumen de la regresión para la variable dependiente: JOB_PROF						
R= .89706447 R <sup>2</sup> = .80472467 Ajustada R <sup>2</sup> = .79623444						
F(1,23)=94.782 p<.00000 Desviación Estándar de la estimación: 8.7676						
N=25	Beta	Desv. Est. de Beta	B	Desv. Est. de B	t(23)	p-level
Intersección			-106.133	20.44719	-5.19058	0.000029
TEST3	0.897064	0.092142	1.968	0.20210	9.73563	0.000000

Al realizar la regresión con una variable se tiene un coeficiente de determinación del 80.47%, por lo que este parámetro disminuyo al original con 4 covariables. La variable *TESTS 3* se mantiene como significativa (dado el nivel de significancia descriptivo o p-level).

**Intervalos de confianza para la intersección y la pendiente.** Se tiene que con una confianza del 95% el valor real de la intersección y de la pendiente están dentro de los intervalos (-149.204, -62.607) y (1.538, 2.394), respectivamente. Al igual que en el ejemplo anterior en ninguno de los dos intervalos está contenido el cero, por lo que se puede decir que ninguno de los dos coeficientes es nulo.

Tabla 3.2.7

Análisis de Varianza; Var. Dep: JOB_PROF					
Fuente de variación	Suma de Cuadrados	Grados de libertad	Cuadrados Medios	F	p-level
Regresión	7285.977	1	7285.977	94.78242	0.000000
Residuales	1768.023	23	76.871		
Total	9054.000				

Se puede advertir en la ANOVA que el modelo matemático para esta variable es considerado como muy bueno dada la significancia de la prueba *F* y en el p-level.

La ecuación de regresión hasta este momento quedaría de la siguiente forma:

$$Y = -106.133 + 1.968X_3$$

Lo cual significa que por cada unidad de cambio de la covariable *TEST3* se tendrá que la variable dependiente aumenta. A la ordenada al origen también se le conoce como estado basal del modelo.

## Análisis de los residuales.

Tabla 3.2.8

						Residuales crudos (Job_prof)		
						Variable dependiente: JOB_PROF		
Caso	Residuales crudos					Valore	Valores	Residuales
	-3s	.	.	0	.	Observados	Predecidos	
1	.	.	.	*	.	88.0000	90.6259	-2.6259
2	.	.	*	.	.	80.0000	88.6583	-8.6583
3	.	.	.	*	.	96.0000	96.5287	-0.5287
4	.	.	.	*	.	76.0000	76.8528	-0.8528
5	.	.	.	*	.	80.0000	80.7880	-0.7880
6	.	.	.	*	.	73.0000	80.7880	-7.7880
7	.	.	.	.	*	58.0000	51.2742	6.7258
8	.	.	.	*	.	116.0000	122.1073	-6.1073
9	.	.	.	.	*	104.0000	102.4315	1.5685
10	.	.	.	*	.	99.0000	100.4639	-1.4639

No existe una evidencia de que se deban quitar observaciones pues ningún punto queda fuera de los límites de sigma que es de  $\pm 3$ , aunque se hará otro análisis para saber si existen “outliers”.

## Distancias de Mahalanobis y los residuales borrados.

Tabla 3.2.10

							Distancias de Mahalanobis (Job_prof)		
							Variable dependiente: JOB_PROF		
Caso	Distancias de Mahalanobis						Valores	Valores	Residuales
	.008	.	.	.	.	5.52	Observados	Predecidos	
1	.	*	.	.	.	.	88.0000	90.6259	-2.6259
2	.	*	.	.	.	.	80.0000	88.6583	-8.6583
3	.	*	.	.	.	.	96.0000	96.5287	-0.5287
4	.	.	*	.	.	.	76.0000	76.8528	-0.8528
5	.	.	*	.	.	.	80.0000	80.7880	-0.7880
6	.	.	*	.	.	.	73.0000	80.7880	-7.7880
7	.	.	.	.	.	*	58.0000	51.2742	6.7258
8	.	.	.	.	*	.	116.0000	122.1073	-6.1073
9	.	.	*	.	.	.	104.0000	102.4315	1.5685
10	.	.	*	.	.	.	99.0000	100.4639	-1.4639

Según el gráfico la observación que más se aleja es la 7, pero se hace la prueba para saber si realmente este es un “outlier” o no lo es.

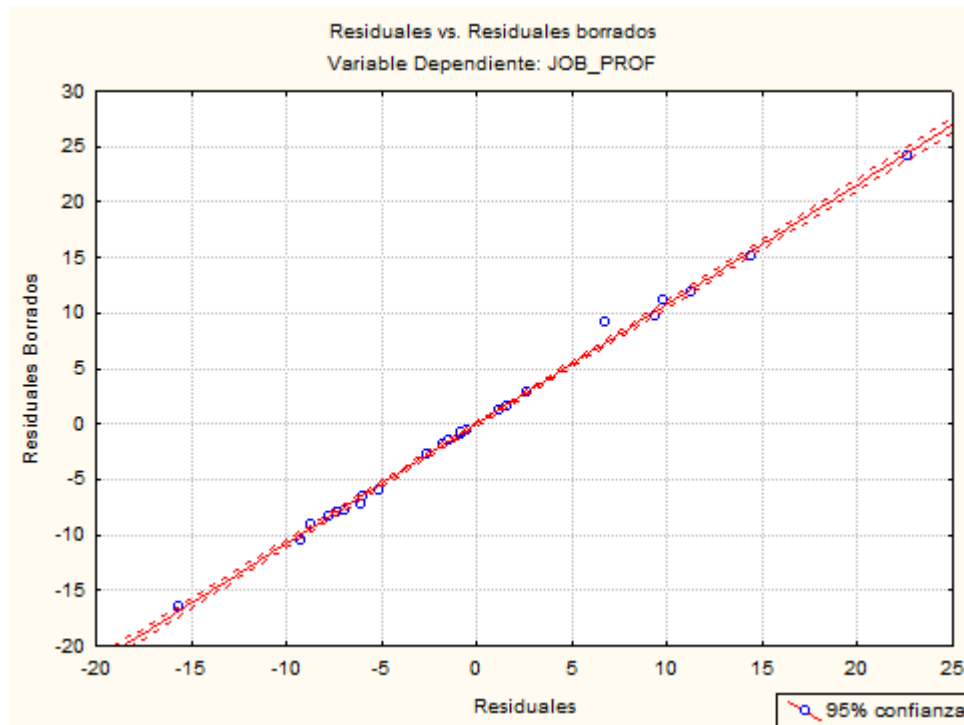
Tabla 3.2.10

Residuales Estandarizados							Residuales Estandarizados: JOB_PROF Outliers			
Caso	-5.	-4.	-3.	±2.	3.	4.	5.	Valores Observados	Valores Predecidos	Residuales
18	.	.	.	.	*	.	.	127.0000	104.3990	22.60096
Mínimo	.	.	.	.	*	.	.	127.0000	104.3990	22.60096
Máximo	.	.	.	.	*	.	.	127.0000	104.3990	22.60096
Media	.	.	.	.	*	.	.	127.0000	104.3990	22.60096
Mediana	.	.	.	.	*	.	.	127.0000	104.3990	22.60096

El punto atípico que resultó en la prueba fue la observación número 18, pues en realidad este punto es el máximo de todas las observaciones (de la variable respuesta) y sale del rango en una cantidad considerable.

La siguiente gráfica es la de los residuales vs residuales borrados

Gráfica 3.2.4

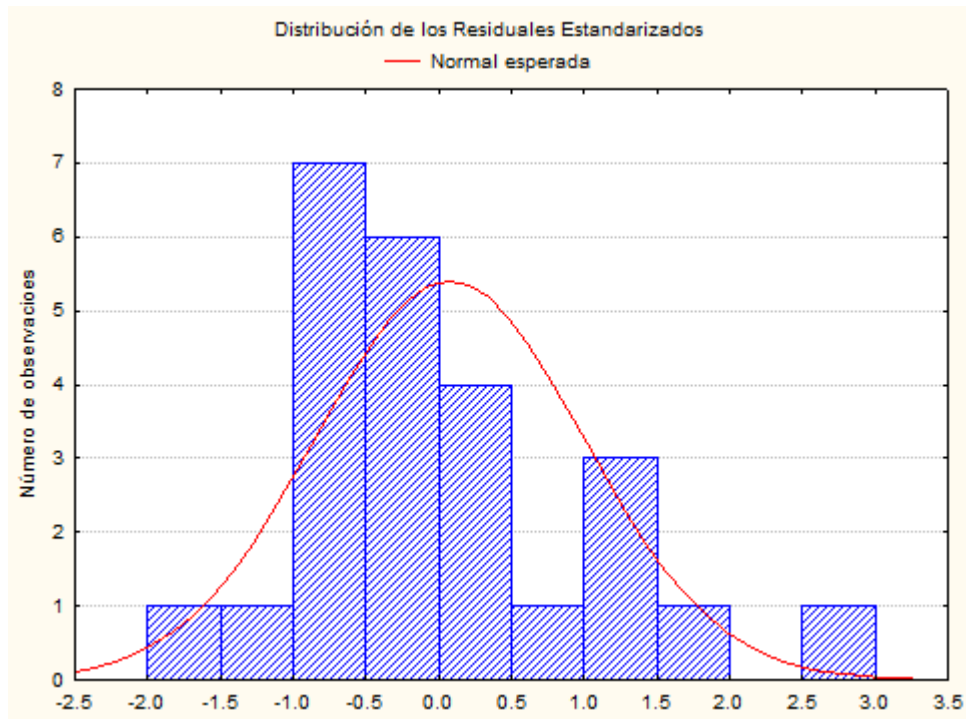


Se observa que la mayoría de los puntos están sobre la recta, por lo que estos residuos son muy parecidos.



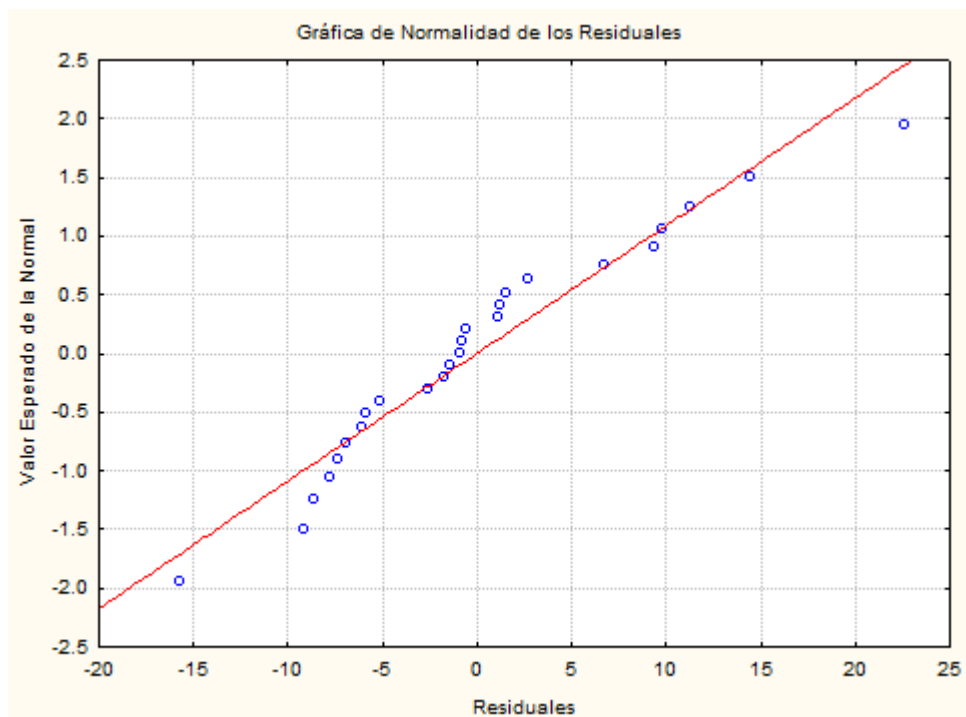
## Distribución de los residuales.

Gráfica 3.2.5



No se puede decir que se tiene una distribución normal de los residuales estandarizados, pues se alejan del cero, que es la media, además se tiene una varianza grande.

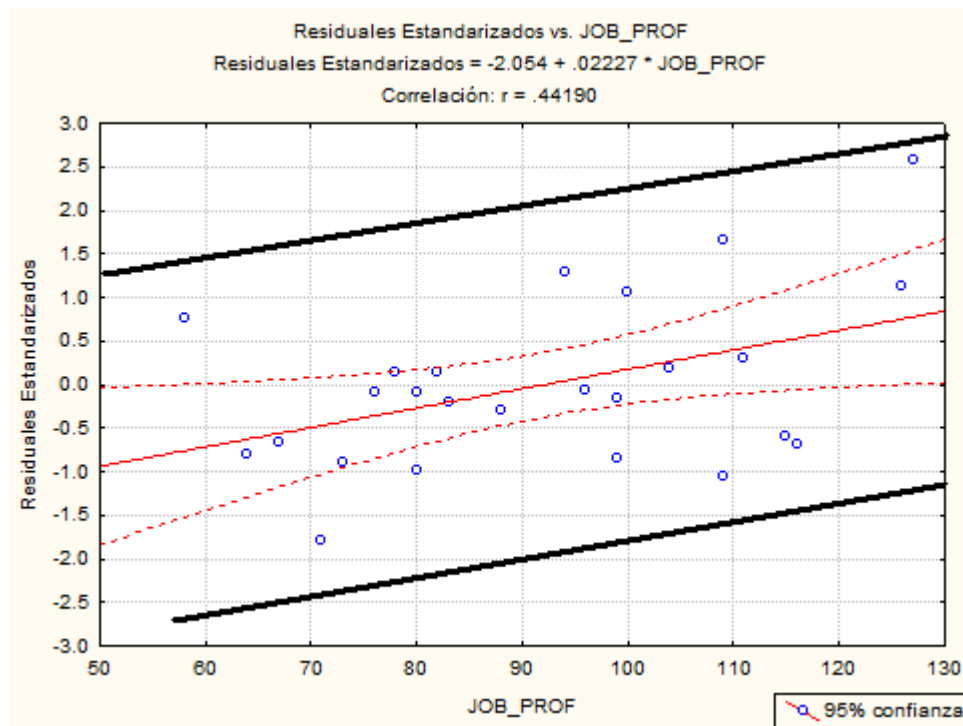
Gráfica 3.2.6



La gráfica de normalidad de los residuos dice que se alejan un poco de la recta, sobre todo en los extremos de esta línea.

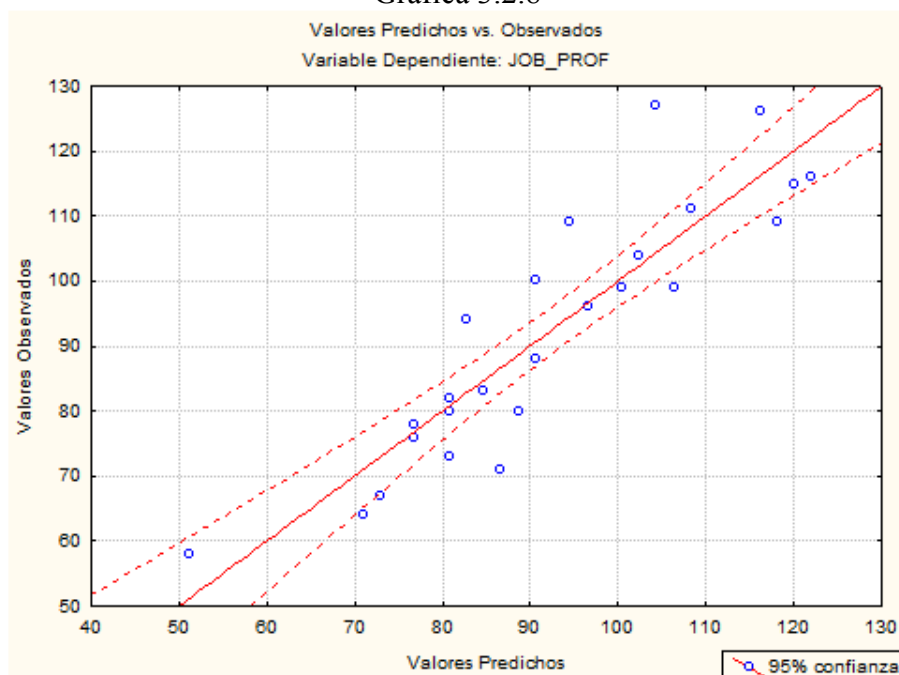
**Varianza Constante.** En el siguiente gráfico, se ve que existe una varianza constante

Gráfica 3.2.7



En la gráfica de los valores predichos vs los valores observados, los puntos se dispersan.

Gráfica 3.2.8



### 3.3 Ejemplo 3

El siguiente ejemplo se encuentra en la página 270 del libro M. Bolstad, William. (2007): *Introduction to Bayesian Statistics*. Editorial: Wiley - Interscience. Second Edition.

Una compañía fabrica un nuevo producto alimenticio y se debe controlar la humedad al final de la producción de dicho comestible. Es más barato y por lo tanto preferible medir el nivel en una etapa del proceso en lugar de al final de la fabricación. El estadístico de la empresa, recomienda a los ingenieros ejecutar el comienzo de medición de la humedad en un nivel dentro del proceso de fase ya que puede dar una buena predicción de lo que el nivel de humedad final será, para lo cual organiza la recolección de los datos a partir de 25 lotes, dado el nivel de humedad dentro del proceso y el nivel de humedad al final de cada lote. Estos se muestran en la siguiente tabla:

Tabla 3.3.1

	En nivel de proceso y en nivel final de humedad			
	1 Nivel de Proceso	2 Nivel Final		
Lote 1	14.36	13.84		
Lote 2	14.48	14.41		
Lote 3	14.53	14.22		
Lote 4	14.52	14.63		
Lote 5	14.35	13.95		
Lote 6	14.31	14.37		
Lote 7	14.44	14.41		
Lote 8	14.23	13.99		
Lote 9	14.32	13.89		
Lote 10	14.57	14.59		
Lote 11	14.28	14.32		
Lote 12	14.36	14.31		
Lote 13	14.5	14.43		
Lote 14	14.52	14.44		
Lote 15	14.28	14.14		
Lote 16	14.13	13.9		
Lote 17	14.54	14.37		
Lote 18	14.6	14.34		
Lote 19	14.86	14.78		
Lote 20	14.28	13.76		
Lote 21	14.09	13.85		
Lote 22	14.2	13.89		
Lote 23	14.5	14.22		
Lote 24	14.02	13.8		
Lote 25	14.45	14.67		

### Descripción de las variables.

$X_1$  será el *Nivel\_Proceso* en una cierta fase dentro de todo proceso.

$Y$  será el *Nivel\_Final* de la fase de todo el proceso.

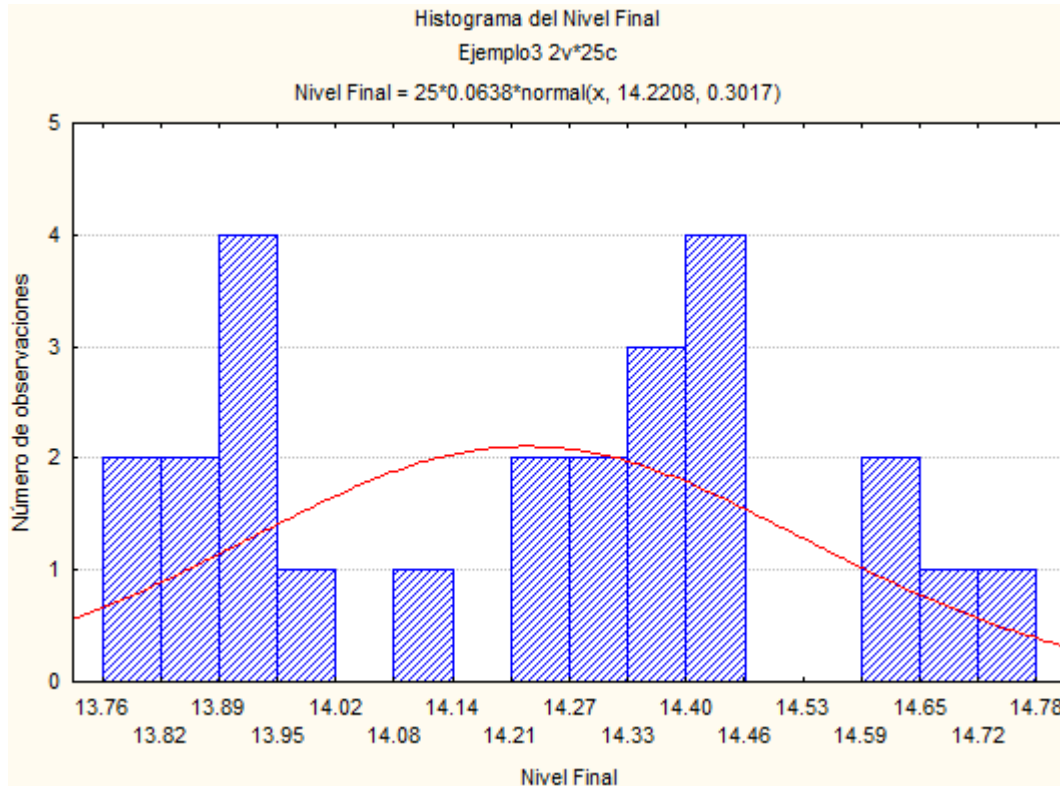
### Distribución de las variables.

Tabla 3.3.2

Variable	Medias y Desv. Est.		
	Medias	Desv. Est.	N
Nivel de Proceso	14.38880	0.184714	25
Nivel Final	14.22080	0.301730	25

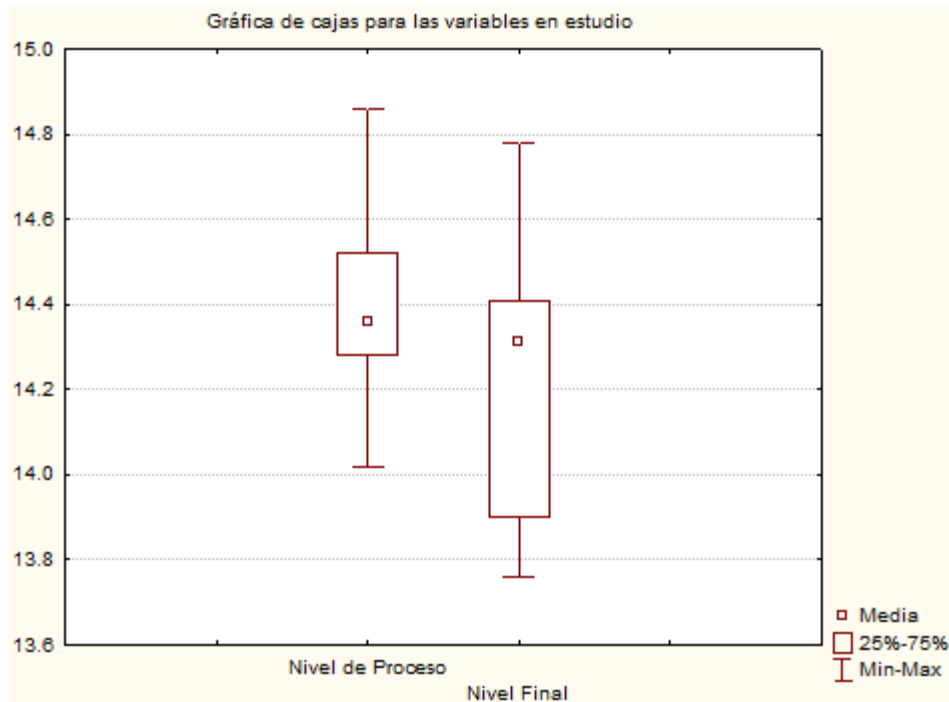
Se pueden observar las medias y las desviaciones estándar de las variables *Nivel de Proceso* y de *Nivel\_Final*.

Gráfica 3.3.1



Es claro ver que no existe una distribución consistente con la normal, de hecho se aleja demasiado de dicha curva. Los datos están bastante dispersos y se alejan de la media.

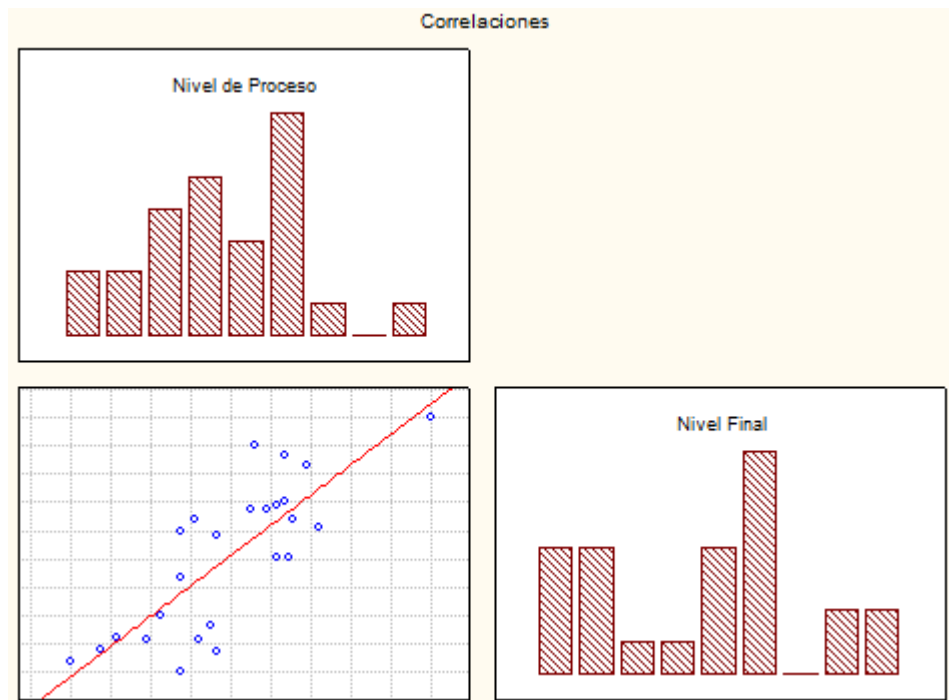
Gráfica 3.3.2



En la gráfica anterior se observa que ambas variables tienen poca dispersión. Las medias son similares, los máximos y mínimos son distintos para cada variable.

En la siguiente gráfica se observan la correlación entre las dos variables.

Gráfica 3.3.3



**Modelo con una variable.** Se realiza el procedimiento ya antes visto en los dos ejemplos anteriores.

Tabla 3.3.3

Resumen de la regresión para la variable dependiente: Nivel Final						
R= .79561512 R²= .63300342 Ajustada R²= .61704705						
F(1,23)=39.671 p<.00000 Desviación Estándar de la estimación: .18672						
N=25	Beta	Desv. Est. de Beta	B	Desv. Est. de B	t(23)	p-level
Intersección			-4.47938	2.969233	-1.50860	0.145016
Nivel de Proceso	0.795615	0.126319	1.29963	0.206341	6.29848	0.000002

Al hacer la regresión lineal simple de este ejemplo, se obtuvo un coeficiente de determinación de 63.3%, además en las pruebas del nivel de significancia de la pendiente y de la intersección (o p-level) se puede observar que para la pendiente es bastante bueno (pues es prácticamente cero), es decir, la variable independiente explica a la respuesta, sin embargo, la intersección resultó como un coeficiente no significativo.

**Intervalos de confianza para la intersección y la pendiente.** Con una confianza del 95% el valor real de la intersección y de la pendientes se encuentra dentro del intervalo (-10.08, 1.51) y (0.88, 1.68), respectivamente. En este caso dentro del primero de los intervalos se encuentra incluido el cero, por lo que se procederá a realizar la prueba de hipótesis correspondiente, para poder determinar si en realidad la intersección es nula.

La prueba es:

$$H_0 : \beta_0 = 0 \text{ vs } H_1 : \beta_0 \neq 0$$

El estadístico a utilizar es:

$$t_0 = \frac{\hat{\beta}_0}{\sqrt{MS_E \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}}$$

Regla de decisión: rechaza la hipótesis nula si  $|t_0| > t_{1-\alpha/2, n-2}$ .

Los valores de cada uno de los términos de la ecuación anterior son los siguientes:

$$n = 25, \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = -4.7938, \bar{x}^2 = \left( \frac{\sum_{i=1}^n x_i}{n} \right)^2 = 207.072,$$

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = 0.7718, SS_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = 0.80188,$$

$$MS_E = \frac{SS_E}{n-2} = 0.03486.$$

Entonces

$$t_0 = \frac{\hat{\beta}_0}{\sqrt{MS_E (1/n + \bar{x}^2/S_{xx})}} = \frac{-4.7938}{\sqrt{(0.03486)(1/25 + 207.072/0.7718)}} = -0.5124.$$

Ahora, tomando el valor absoluto de  $t_0$  se obtiene que  $|t_0| = 0.5124$ , y buscando el valor en tablas del estadístico *t de Student* se tiene lo siguiente,  $t_{1-\alpha/2, n-2} = t_{95\%, 23} = 1.3195$ .

Comparando ambos resultados se tiene que  $|t_0| < t_{1-\alpha/2, n-2}$ , y por lo tanto no se rechaza la hipótesis nula, es decir, la recta pasa por el origen.

Se corre nuevamente una regresión pero ahora con la intersección igual a cero.

Tabla 3.3.4

Resumen de la regresión para la variable dependiente: Nivel Final						
R= .99991288 R²= .99982577 Ajustada R²= .99981851						
F(1,24)=1377E2 p<0.0000 Desviación Estándar de la estimación: .19162						
	Beta	Desv. Est. de Beta of Beta	B	Desv. Est. de B of B	t(24)	p-level
N=25						
<b>Nivel de Proceso</b>	<b>0.999913</b>	<b>0.002694</b>	<b>0.988373</b>	<b>0.002663</b>	<b>371.1174</b>	<b>0.00</b>

Se obtuvo un coeficiente de determinación de 63.3%. Hay que hacer notar que el  $R^2$  que aparece en la tabla igual a 99.98% no es el coeficiente de determinación verdadero<sup>13</sup>.

<sup>13</sup> En una regresión a la que se le quita la intersección, la  $R^2$  (coeficiente de determinación) representa el porcentaje de explicación de la variabilidad en el origen, no es lo mismo que la  $R^2$  cuando sí se incluye al origen.

En la ANOVA se tiene:

Tabla 3.3.5

Análisis de Varianza; Var. Dep: Nivel Final (Ejemplo3)					
Fuente de variación	Suma de Cuadrados	Grados de libertad	Cuadrados Medios	F	p-level
Regresión	5057.083	1	5057.083	137728.1	0.00
Residuales	0.881	24	0.037		
Total	5057.964				

Los cuadrados medios asociados a la regresión tienen un coeficiente muy alto comparados con los cuadrados medios de los residuales, por lo que el estadístico  $F$  señala que la fuente de variación debida a la regresión es 137728.1 veces más grande que la de los errores, además el nivel de significancia de la prueba indica que se trata de una excelente regresión.

La ecuación de regresión simple quedaría de la siguiente forma:

$$Y = 0.988373X_1$$

Significa que por cada cambio en una unidad de la variable independiente la dependiente aumenta en 0.988373 unidades.

La tabla de correlaciones dice lo siguiente:

Variable	Correlaciones (Ejemplo3)	
	Nivel de Proceso	Nivel Final
Nivel de Proceso	1.000000	0.795615
Nivel Final	0.795615	1.000000

Se observa que preexiste una correlación lineal de 0.795615 entre la variable explicativa y la de respuesta.



**Análisis de los residuales.** Se examinarán los residuos con lo gráficos de éstos.

**Gráficas de los residuos.**

Tabla 3.3.7

Caso	Residuales crudos						Residual crudos (Ejemplo3) Variable Dependiente: Nivel Final		
	-3s	.	.	0	.	+3s	Valores Observados	Valores Predecidos	Residuales
	Lote 1	.	*	.	.	.	.	13.84000	14.19304
Lote 2	.	.	.	.	*	.	14.41000	14.31165	0.098351
Lote 3	.	.	*	.	.	.	14.22000	14.36107	-0.141067
Lote 4	.	.	.	.	.	*	14.63000	14.35118	0.278817
Lote 5	.	.	*	.	.	.	13.95000	14.18316	-0.233160
Lote 6	.	.	.	.	.	*	14.37000	14.14362	0.226376
Lote 7	.	.	.	.	*	.	14.41000	14.27211	0.137887
Lote 8	.	.	.	*	.	.	13.99000	14.06456	-0.074555
Lote 9	.	.	*	.	.	.	13.89000	14.15351	-0.263508
Lote 10	.	.	.	.	*	.	14.59000	14.40060	0.189399

En este ejemplo no hay casos que queden fuera de los límites de sigma de  $\pm 3$ , con lo cual se puede pensar que no se hallarán “outliers”.

**Gráfica de “outliers”.** Cuando uno presiona el botón *Casewise plot of outliers* en la pantalla aparece el mensaje de que no existen puntos atípicos (pues están dentro del límite de  $\pm 2$  de la desviación estándar).

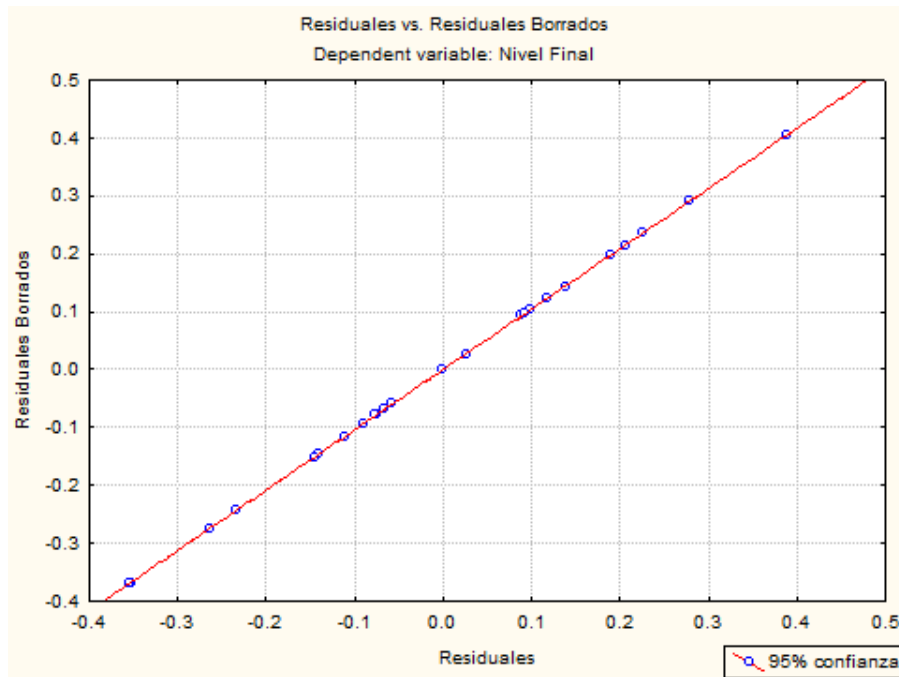
**Distancias de Mahalanobis y los residuales borrados.**

Tabla 3.3.8

Case	Distancias de Mahalanobis						Distancias de Mahalanobis (Ejemplo3) Variable Dependiente: Nivel Final		
	.949	.	.	.	.	1.07	Valores Observados	Valores Predecidos	Residuales
	Lote 1	.	.	*	.	.	.	13.84000	14.19304
Lote 2	.	.	.	.	*	.	14.41000	14.31165	0.098351
Lote 3	.	.	.	.	*	.	14.22000	14.36107	-0.141067
Lote 4	.	.	.	.	*	.	14.63000	14.35118	0.278817
Lote 5	.	.	*	.	.	.	13.95000	14.18316	-0.233160
Lote 6	.	.	*	.	.	.	14.37000	14.14362	0.226376
Lote 7	.	.	.	*	.	.	14.41000	14.27211	0.137887
Lote 8	.	.	*	.	.	.	13.99000	14.06456	-0.074555
Lote 9	.	.	*	.	.	.	13.89000	14.15351	-0.263508
Lote 10	.	.	.	.	*	.	14.59000	14.40060	0.189399

Ninguna observación queda fuera de los límites, pues todos presentan distancias muy pequeñas. Al observar los residuales se distingue que éstos son de casi cero.

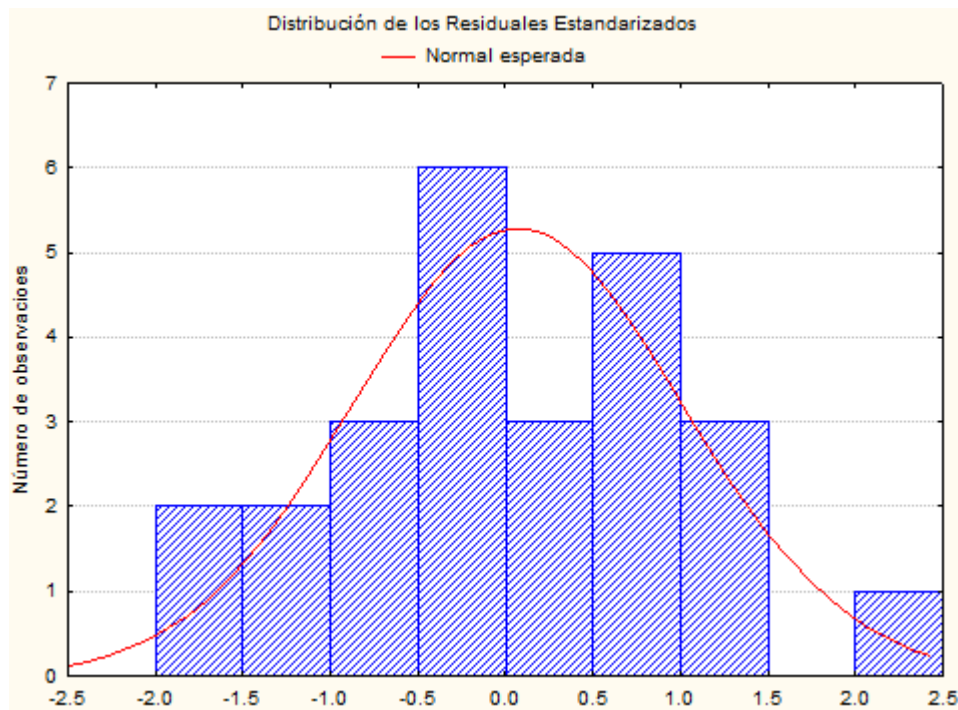
Gráfica 3.3.4



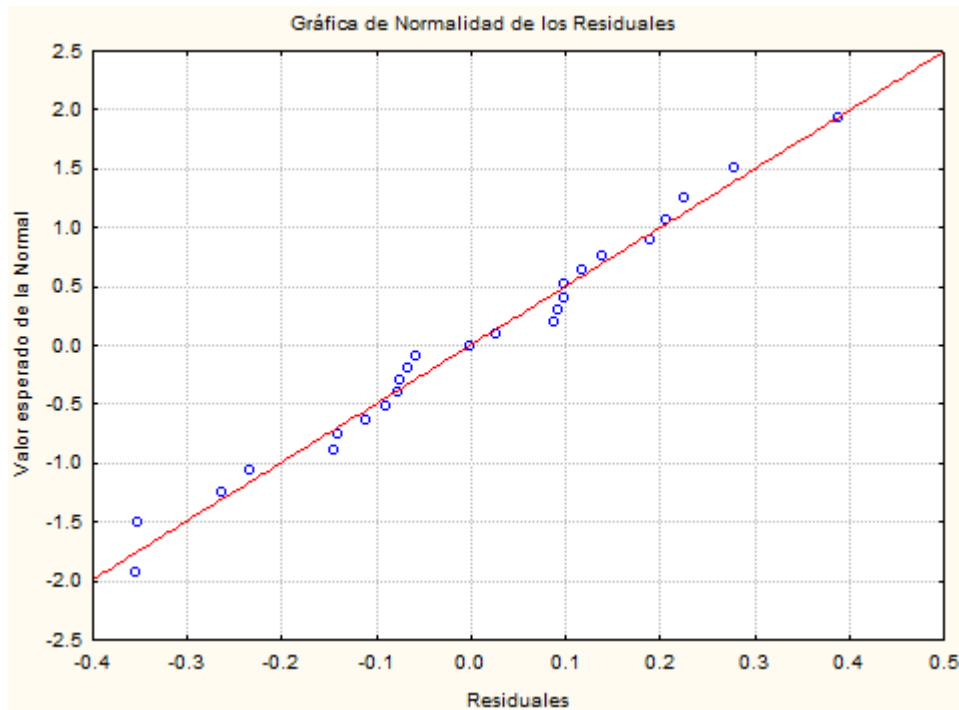
Es más sencillo observar en esta gráfica que no existen “outliers”.

**Distribución de los residuales.** La gráfica puede sugerir que los residuos se acercan a la normalidad, sólo una columna sale de la distribución en un rango (la de la derecha).

Gráfica 3.3.5



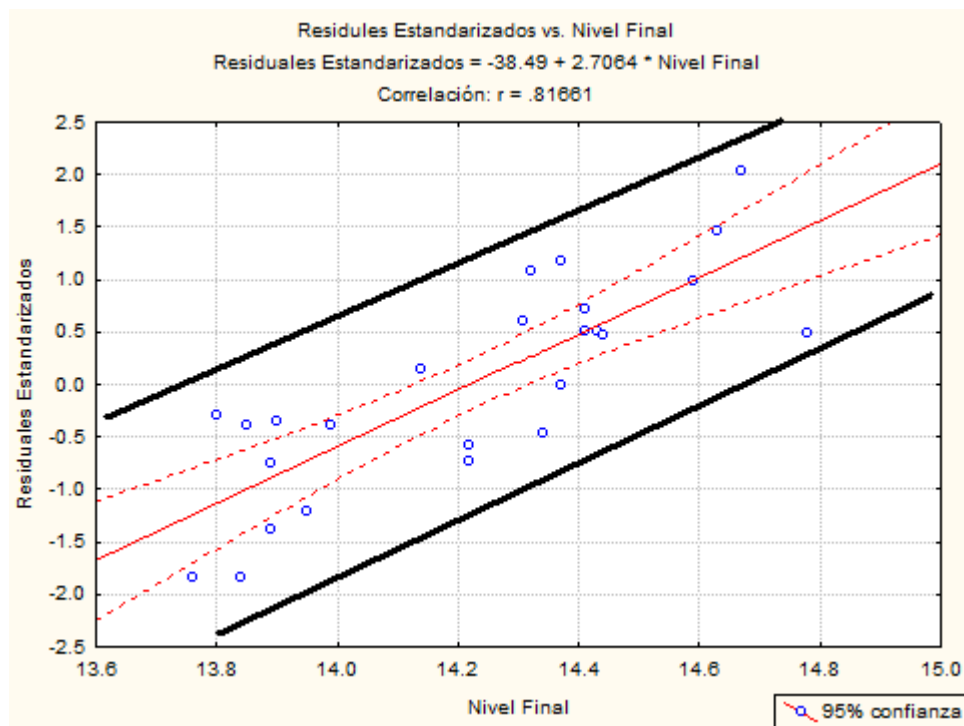
Gráfica 3.3.6



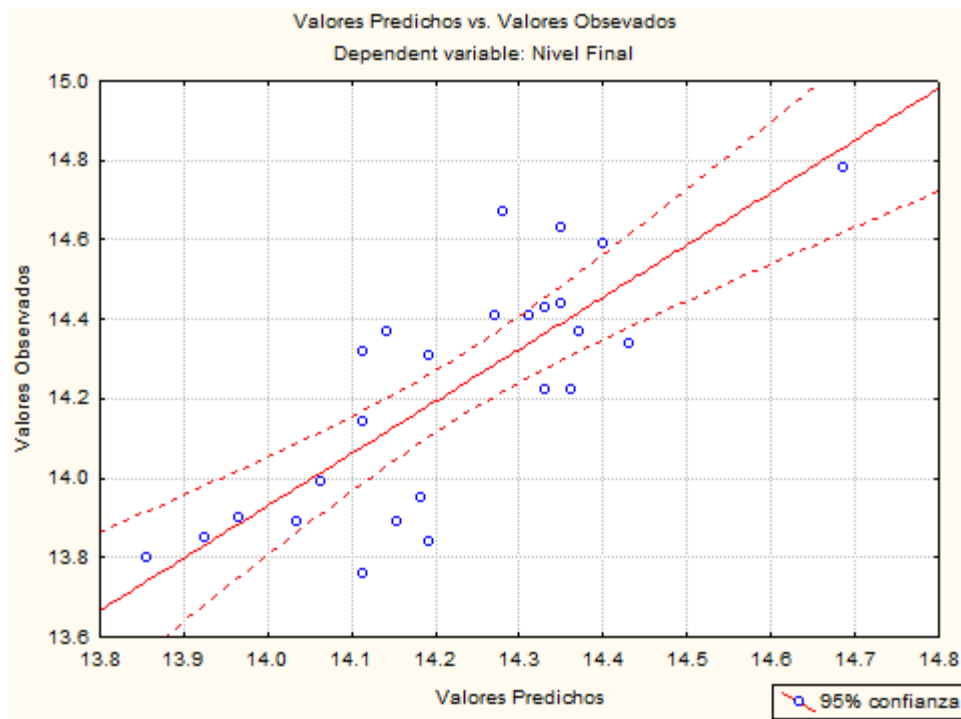
Se puede notar que parecen seguir la línea recta de la gráfica, con esto se podría creer que se distribuyen normalmente, si no fuera el caso, entonces se desviarían de la recta.

**Varianza Constante.** En el siguiente gráfico, se ve que existe una varianza constante.

Gráfica 3.3.7



Gráfica 3.3.8



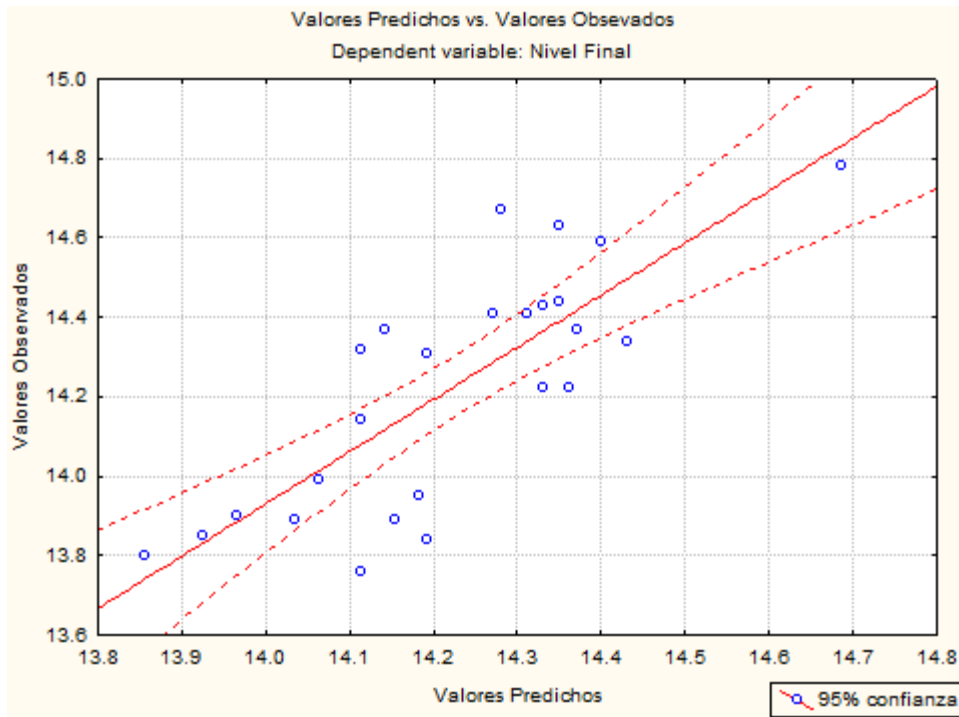
En esta última gráfica se aprecia que hay puntos que se acercan y que se alejan de la recta, con lo cual se puede decir que el modelo planteado no es tan malo

#### 4ta Parte: Ejemplos de regresión lineal simple con enfoque bayesiano

Para la aplicación de los ejemplos, se utilizará el programa conocido como **R**, que es un sistema para cómputo estadístico y gráficos de licencia libre que se encuentra en Internet, el cual consiste de un lenguaje de programación, un entorno de ejecución, un debugger y la habilidad de correr programas guardados en archivos de tipo script. Su diseño fue influenciado por dos lenguajes existentes: S y Scheme. Como resultado, el lenguaje R es muy similar en apariencia al de S, con la semántica similar a la de Scheme.

El paquete para **R** que se manejará es el denominado **Bolstad**, pues opera con una programación para el análisis de regresión lineal simple bayesiana. El comando principal para el análisis requerido es el **bayes.lin.reg**.

Gráfica 3.3.8



En esta última gráfica se aprecia que hay puntos que se acercan y que se alejan de la recta, con lo cual se puede decir que el modelo planteado no es tan malo

#### 4ta Parte: Ejemplos de regresión lineal simple con enfoque bayesiano

Para la aplicación de los ejemplos, se utilizará el programa conocido como **R**, que es un sistema para cómputo estadístico y gráficos de licencia libre que se encuentra en Internet, el cual consiste de un lenguaje de programación, un entorno de ejecución, un debugger y la habilidad de correr programas guardados en archivos de tipo script. Su diseño fue influenciado por dos lenguajes existentes: S y Scheme. Como resultado, el lenguaje R es muy similar en apariencia al de S, con la semántica similar a la de Scheme.

El paquete para **R** que se manejará es el denominado **Bolstad**, pues opera con una programación para el análisis de regresión lineal simple bayesiana. El comando principal para el análisis requerido es el **bayes.lin.reg**.

## Descripción del paquete Bolstad:

La función **bayes.lin.reg.** se utiliza para encontrar la distribución *posteriori* de las variables de regresión lineal simple, es decir, para la pendiente y para la intersección, cuando se tiene una muestra de los pares ordenados  $(x_i, y_i)$  del modelo de regresión lineal simple

$$y_i = \alpha_x + \beta x_i + \varepsilon_i,$$

donde los errores de las observaciones son los  $\varepsilon_i$ , independientes y con distribución  $normal(0, \sigma^2)$ .

El programa que se utilizará es:

**bayes.lin.reg**(y, x, slope.prior = “no informativa”, intcpt.prior = “no informativa”, mb0 = 0, sb0 = 0, ma0 = 0, sa0 = 0, sigma = NULL, alpha = 0.05, plot.data = FALSE, pred.x = NULL, ret = FALSE)

Donde:

y es el vector de las respuestas.

x es el valor de la variable explicativa asociada con cada respuesta.

**slope.prior** se utiliza un “flat”<sup>14</sup> o una “normal” *a priori* para la pendiente  $\beta$ .

**intcpt.prior** se utiliza un “flat” o una “normal” *a priori* para la intersección  $\alpha_x$ .

**mb0** es la media *a priori* de la regresión lineal simple para  $\beta$ . Este argumento es ignorado para los “flat” *a priori*.

**sb0** es el valor de la desviación estándar de la regresión lineal simple para  $\beta$ , éste debe ser mayor que cero. El argumento es ignorado para los “flat” *a priori*.

**ma0** es la media *a priori* de la regresión lineal simple para  $\alpha_x$ . Este argumento es ignorado para los “flat” *a priori*.

---

<sup>14</sup> En inglés se utiliza la palabra *flat a priori* para denotar a las distribuciones *a priori* no informativas

**sa0** es el valor de la desviación estándar de la regresión lineal simple para  $\alpha_x$ , éste debe ser mayor que cero. Este argumento es ignorado para los “flat” *a priori*.

**sigma** es el valor de la desviación estándar de los residuos. Por omisión, se supone que se desconoce y el valor calculado de la muestra se utiliza en su lugar. Esto afecta cuando se obtienen los intervalos de credibilidad.

**alpha** es lo que controla el ancho del intervalo de credibilidad.

**plot.data** si es cierto entonces los datos se representan gráficamente y la línea de regresión *posteriori* superpuesta a la datos.

**pred.x** es un vector de valores de  $x$  para predecir valores de  $y$  y la desviación estándar de los errores de la predicción. Si es nulo entonces esto no será devuelto. Si es verdadero entonces devuelve **pred.y** que es el vector de los valores predichos correspondientes a **pred.x**, y además será devuelto **pred.se** que son los errores estandarizados de los valores predichos de **pred.y**.

**ret** cuando es cierto, entonces la *a priori* y la *a posterior* se devuelven como una lista.

## 4.1 Ejemplo 1

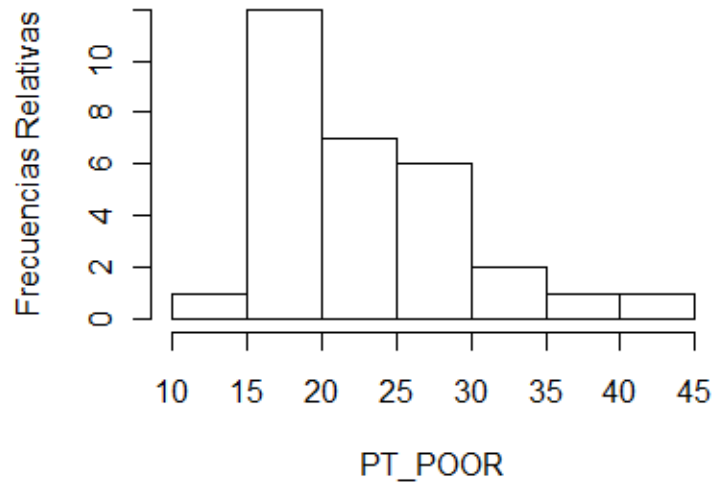
Los datos son los mismos que se consideraron en el primer ejemplo de la tercera parte.

En el programa de R existen comandos que ayudan al análisis estadístico. Por ejemplo, para hacer un histograma se podría utilizar el comando:

```
hist(y,freq=FALSE,main="Histograma",xlab="PT_POOR",ylab="Frecuencias  
Relativas")
```

Gráfica 4.1.1

### Histograma

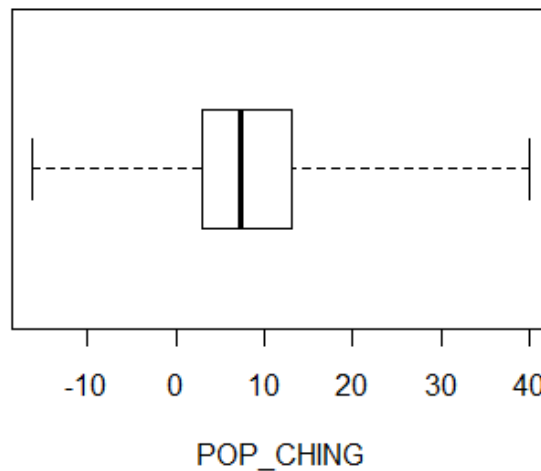


El siguiente comando sirve para obtener una gráfica de cajas de *POP\_CHING*

```
boxplot(x,range=0,main="DiagramadeCajas",xlab="POP_CHING",horizontal= TRUE)
```

Gráfica 4.1.2

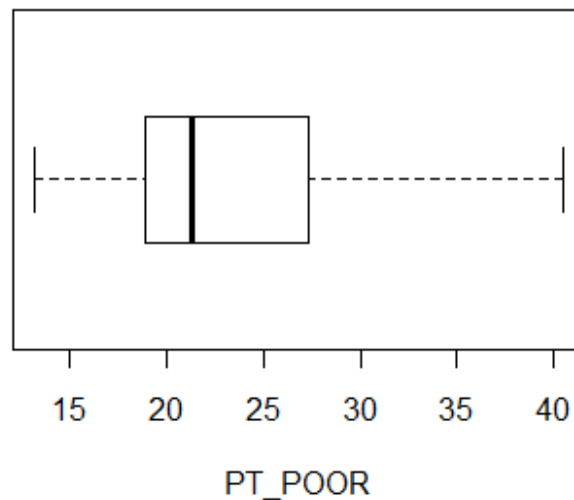
### Diagrama de Cajas



```
boxplot(y,range=0,main="Diagrama de Cajas",xlab="PT_POOR",horizontal=TRUE)
```



Gráfica 4.1.3

**Diagrama de Cajas**

Ahora, se correrá un programa que contenga las variables  $PT\_POOR$  como la de respuesta y a  $POP\_CHING$  como la explicativa. Para el análisis se calculan los datos necesarios:

$$\bar{x} = 7.8670, \bar{y} = 23.0100, \overline{x^2} = 165.0869, \overline{xy} = 139.3547.$$

Hay que hacer notar que es necesario mantener todas las cifras significativas en los términos cuadráticos. La fórmula de  $\beta_1$  utiliza subtracciones, y si se redondea demasiado pronto las diferencias tendrán muy pocas cifras significativas y la precisión se perderá.

$$B = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{139.3547 - 7.867 \times 23.01}{165.0869 - (7.867)^2} = \frac{-41.665}{103.1972} = -0.4037.$$

La recta por mínimos cuadrados es:

$$y = \bar{y} + \beta_1(x - \bar{x}) = 23.01 - 0.4037 \times (x - 7.867).$$

Habrá que decidir que función *a priori* se utilizará.

Se eligió la  $normal(-0.4, (0.1)^2)$  para  $\beta_1$ , pues dado el estimador  $B$  se puede escoger un coeficiente para la media cercano a -0.4037. Al prestar atención a los datos es posible decidir que la media para  $\alpha_x$  podría ser igual a 23. Además, el valor de la diferencia del límite superior menos el límite inferior de los valores de  $y$  es 27.3, y al dividir esta cantidad entre 4 resulta 6.825. Por lo que, se tomará una desviación estándar de 7. Por lo tanto se tiene una *a priori normal*(23, (7)<sup>2</sup>) para  $\alpha_x$ .

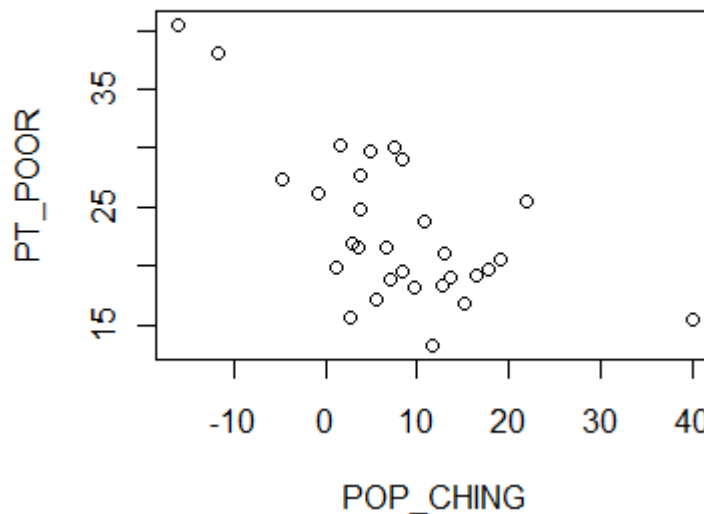
Se corre el siguiente programa para ver la dispersión de los datos.

```
x<-c(13.7,-0.8,9.6,40,8.41,3.5,3,7.1,13,10.7,-16.2,6.6,21.9,17.8,-11.8,7.5,3.7,1.6,8.4,2.7,5.6,
12.7,-4.8,16.5,15.2,11.6,4.9,1.1,3.8,19)
y<-c(19,26.2,18.1,15.4,29,21.6,21.9,18.9,21.1,23.8,40.5,21.6,25.4,19.7,38,30.1,24.8,30.3,19.5,
15.6,17.2,18.4,27.3,19.2,16.8,13.2,29.7,19.8,27.7,20.5)
plot(x,y,main="Diagrama de Dispersion",xlab="POP_CHING",ylab="PT_POOR")
```

El cual da el siguiente gráfico:

Gráfica 4.1.4

### Diagrama de Dispersión



También es posible obtener el coeficiente de correlación con el comando `cor(x,y,method="pearson")`, lo que resulta es -0.6491112, en este caso es una correlación negativa entre las variables.

Luego se corre el programa para la regresión bayesiana con una distribución normal tanto para la pendiente como para la intersección.

El programa es:

```
bayes.lin.reg(y, x, slope.prior = "normal", intcpt.prior = "normal", mb0 = -0.4, sb0 = 0.1, ma0 = 23, sa0 = 7, sigma = NULL, alpha = 0.05, plot.data = TRUE, pred.x = NULL, ret = FALSE)
```

El cual arroja la siguiente salida:

Tabla 4.1.1

Desviación estándar de los residuos: 4.98

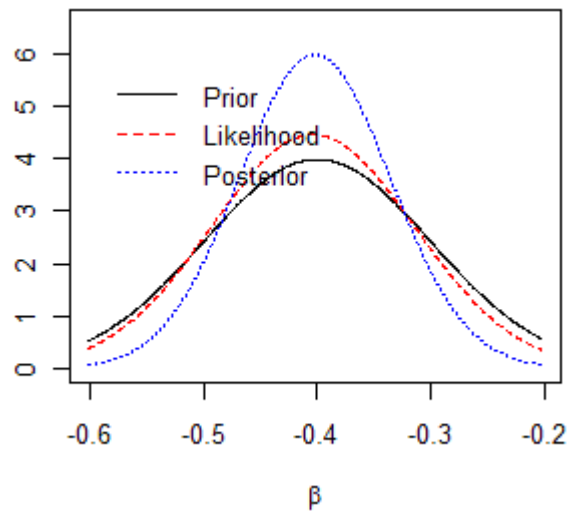
	Media Posteriori	Desv. Est. Posteriori
	-----	-----
Intersección:	23.01	0.90079
Pendiente:	-0.4021	0.066656

Aquí, se puede ver que los parámetros para la distribución normal *posteriori* de la intersección con media y la desviación estándar son 23.01 y 0.90079, respectivamente; los parámetros para la distribución normal *posteriori* para la pendiente son -0.4021 y 0.066656, cada uno de ellos para la media y la desviación estándar. Además la desviación estándar estimada de los residuos es igual a 4.98.

Las distribuciones *a priori* y *posteriori* de la pendiente y de la intersección se muestran en las siguientes gráficas, así como la distribución de la *función de verosimilitud*.

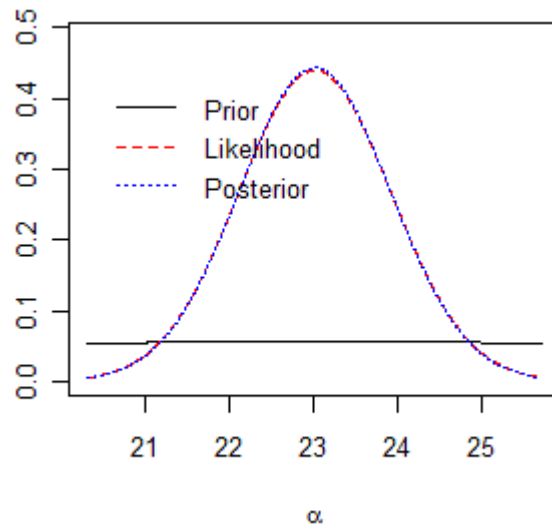
Se puede ver que todas las distribuciones de las funciones para la pendiente, son diferentes cada una entre si. Se observar que la función *posteriori* tiene una varianza mucho menor a las otras dos distribuciones, además todas las curvas están centradas en la media propuesta y que cada una de ellas tiene su varianza muy distintas a las otras.

Gráfica 4.1.5



Distribución *a priori*, *posteriori* y la *función de verosimilitud* de la pendiente

Gráfica 4.1.6



Distribución *a priori*, *posteriori* y la *función de verosimilitud* de la intersección

En la gráfica anterior se aprecia que tanto la distribución *posteriori* como la *función de verosimilitud* son iguales y la *a priori* es una recta horizontal.

Como no se conoce el verdadero valor de la varianza se utilizará la estimada para encontrar un intervalo de credibilidad al 95% para la pendiente con la fórmula  $m'_{\beta_1} \pm t_{1-\alpha/2}^{n-2} \sqrt{(s'_{\beta_1})^2}$ . El intervalo que se encontró fue (-0.53859, -0.26552). Es evidente que en éste no está contenido el cero, claramente se puede rechazar la hipótesis nula

( $H_0 : \beta_1 = 0$  vs  $H_1 : \beta_1 \neq 0$ ). Entonces se establece que para predecir el porcentaje de familias que viven por debajo del nivel de pobreza se puede utilizar el cambio poblacional ocurrido de 1960 a 1970.

Del mismo modo, en R está la opción de correr un programa con funciones *a priori* no informativas tanto para la pendiente como para la intersección:

```
bayes.lin.reg(y, x, slope.prior = "flat", intcpt.prior = "flat", mb0 = -0.4, sb0 = 0.1, ma0 = 23, sa0 = 7, sigma = NULL, alpha = 0.05, plot.data = TRUE, pred.x = NULL, ret = FALSE)
```

La salida es:

Tabla 4.1.2

Desviación estándar de los residuos: 4.98

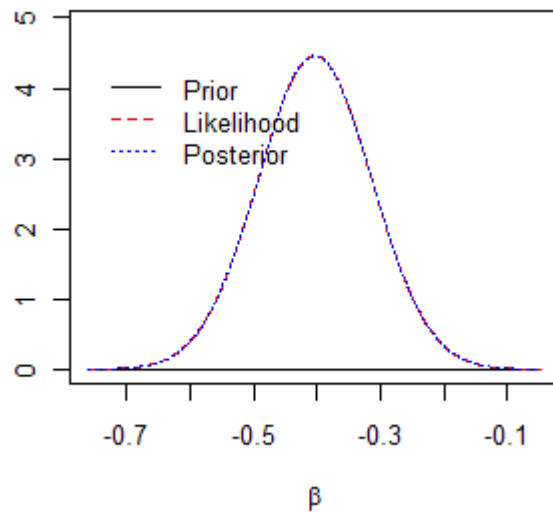
	Media Posteriori	Desv. Est. Posteriori
	-----	-----
Intersección:	23.01	0.90834
Pendiente:	-0.4037	0.089416

La tabla muestra las distribuciones *posteriori* de la pendiente igual a -0.4037 y una intersección equivalente a 23.01, además de los coeficientes de las desviaciones estándares de 0.90834 y 0.089416, respectivamente. Del mismo modo se especifica la desviación estándar de los residuos igual a 4.98. Los gráficos de estas funciones son:

En las gráficas siguientes se aprecia que no hay diferencias entre las tres curvas (la distribución *a priori*, la distribución *posteriori* y la *función de máxima verosimilitud*). Esto es porque la distribución *a priori* no informativa es justamente una función que no aporta ninguna clase de información antes de correr la regresión, y que por lo tanto, la distribución *posteriori* no se ve afectada de alguna forma.

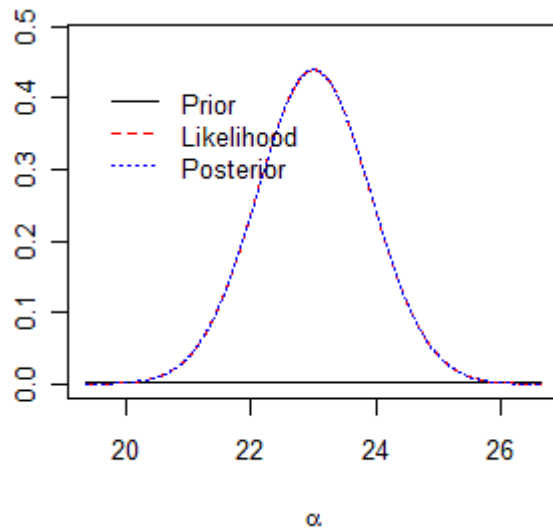
Gráfica para las *a priori* no informativas de la pendiente

Gráfica 4.1.7



Gráfica para las *a priori* no informativas de la intersección.

Gráfica 4.1.8



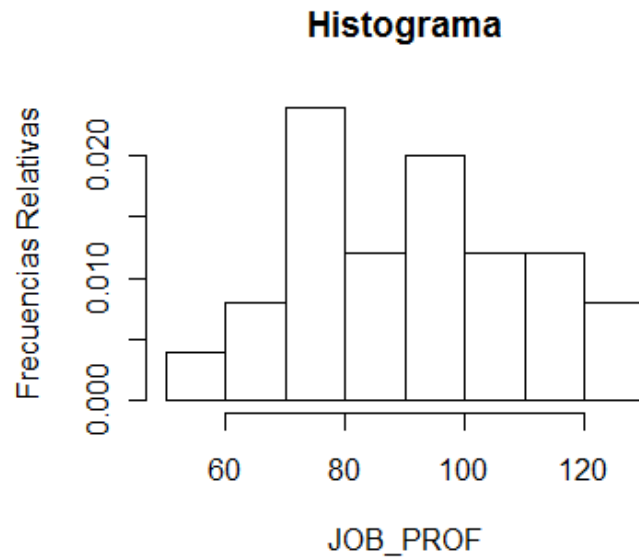
## 4.2 Ejemplo 2

Se utilizan los mismos datos del segundo ejemplo del capítulo 3.

Se harán análisis análogos al ejemplo anterior.

El histograma para *JOB\_PROF* es:

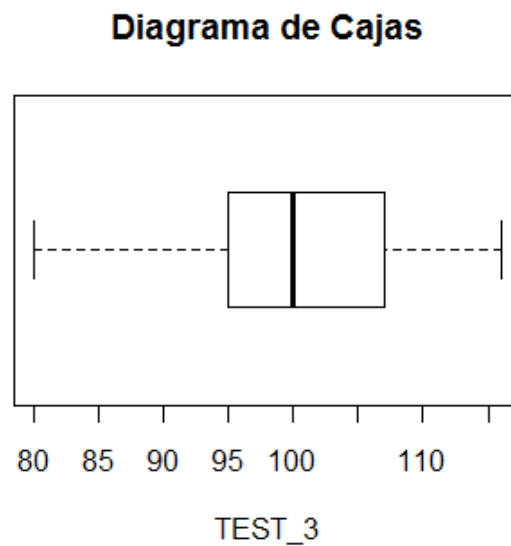
Gráfica 4.2.1



Se observa que la distribución de los datos no se acerca a la normal.

El diagrama de cajas para la variable explicativa es:

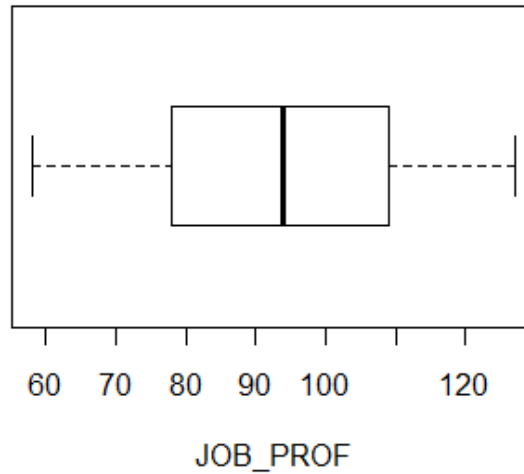
Gráfica 4.2.2



El diagrama de cajas para la variable respuesta es:

Gráfica 4.2.3

### Diagrama de Cajas



Para la obtención del valor de la pendiente estimada se necesitan los siguientes datos:

$$\bar{x} = 100.8, \bar{y} = 92.2, \overline{x^2} = 10235.9, \overline{xy} = 9441.8$$

Entonces, se calcula el estimador de la pendiente como sigue:

$$B = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{9441.8 - 100.8 \times 92.2}{10235.9 - (100.8)^2} = \frac{148.04}{75.26} = 1.967$$

La recta por mínimos cuadrados es

$$y = \bar{y} + \beta_1(x - \bar{x}) = 92.2 + 1.967 \times (x - 100.8)$$

Habrá que decidir que función *a priori* se utilizará.

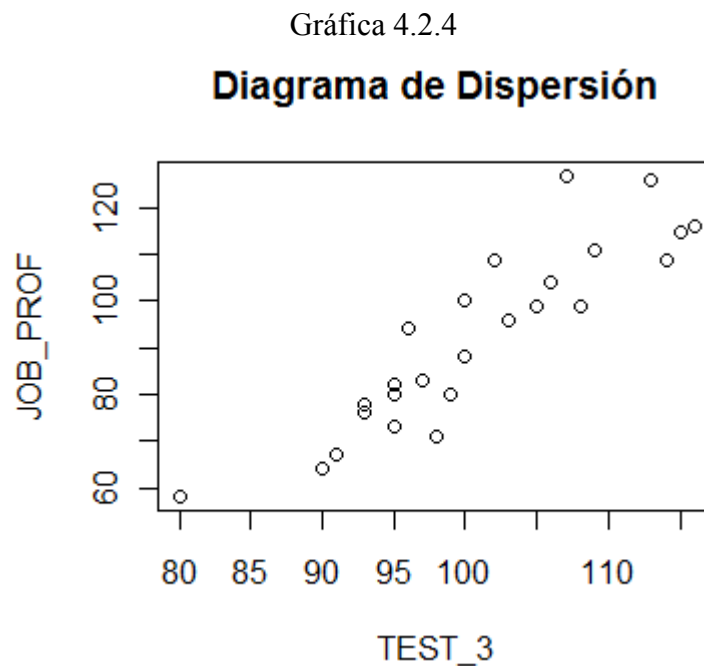
Se eligió la *normal*(2, (0.5)<sup>2</sup>) para  $\beta_1$  y una *a priori normal*(93, (19)<sup>2</sup>) para  $\alpha_x$ .



El programa para observar la grafica de dispersión de los datos es:

```
x<-c(,99,103,93,95,95,80,116,106,105,90,113,96,98,109,102,100,107,108,95,91,114,  
93,115,97)  
y<-c(88,80,96,76,80,73,58,116,104,99,64,126,94,71,111,109,100,127,99,82,67,109,78,  
115,83)  
plot(x,y,main="Diagrama de Dispersion",xlab="TEST_3",ylab="JOB_PROF")
```

El cual da el siguiente gráfico:



El coeficiente de correlación es de 0.8970645.

Asimismo, se corre el programa para la regresión bayesiana con una distribución normal tanto para la pendiente como para la intersección.

```
bayes.lin.reg(y, x, slope.prior = "normal", intcpt.prior = "normal", mb0 = 2, sb0 = 0.5,  
ma0 = 93, sa0 = 19, sigma = NULL, alpha = 0.05, plot.data = TRUE, pred.x = NULL,  
ret = FALSE)
```

El cual manda la siguiente salida:

Tabla 4.2.1

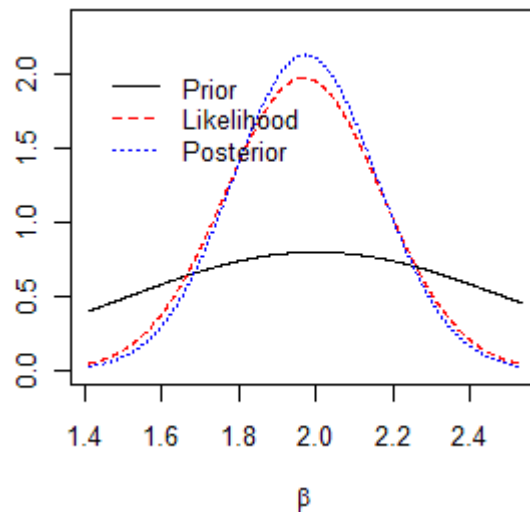
Desviación estándar de los residuos: 8.77

	Media Posteriori	Desv. Est. Posteriori
	-----	-----
Intersección:	92.21	1.7461
Pendiente:	1.972	0.18737

Es muy claro que las distribuciones *posteriori* para la intersección y para la pendiente son:  $normal(92.21, (1.7461)^2)$  y  $normal(1.971, (0.18737)^2)$ , respectivamente. Igualmente se obtuvo un coeficiente para la desviación estándar estimada de los residuos de 8.77.

Al mismo tiempo las siguientes gráficas son expuestas, en las cuales aparece la distribución *a priori* y *posteriori* de la pendiente y la intersección.

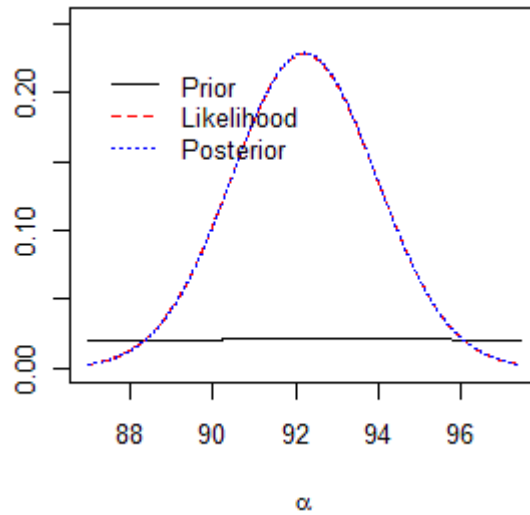
Gráfica 4.2.5



Distribución *a priori*, *posteriori* y la función de verosimilitud de la pendiente

Se observa que las tres curvas son distintas entre sí.

Gráfica 4.2.6



Distribución *a priori*, *posteriori*  
y la *función de verosimilitud* de la intersección

Se puede ver en la gráfica que la distribución *posteriori* y la *función de verosimilitud* son idénticas y la línea recta es la distribución *a priori*.

Un intervalo de credibilidad al 95% para la pendiente es (1.68958, 2.27775). Éste no agrupa al cero, por lo que no tiene sentido probar la hipótesis nula de que este parámetro sea igual a cero, dado que se rechazará. Entonces se concluye que los resultados de la tercera prueba de aptitud pueden utilizarse para predecir el dominio del trabajo después del periodo de prueba.

Ahora, si se hace con la distribución *a priori* no informativa.

**bayes.lin.reg**(y, x, slope.prior = "flat", intcpt.prior = "flat", mb0 = 2, sb0 = 0.5, ma0 = 93, sa0 = 19, sigma = NULL, alpha = 0.05, plot.data = TRUE, pred.x = NULL, ret = FALSE)

Resulta lo siguiente:

Tabla 4.2.2

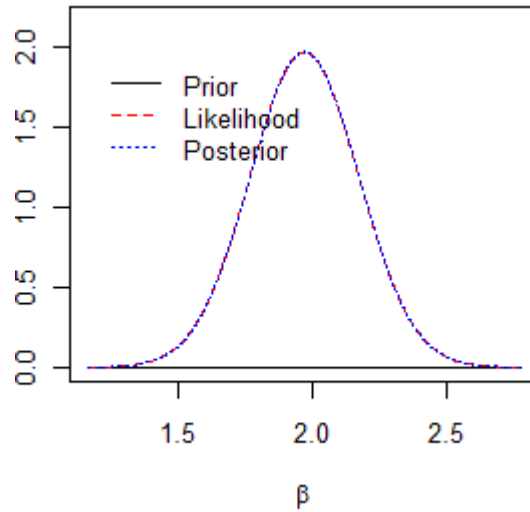
Desviación estándar de los residuos: 8.77

Media Posteriori    Desv. Est. Posteriori

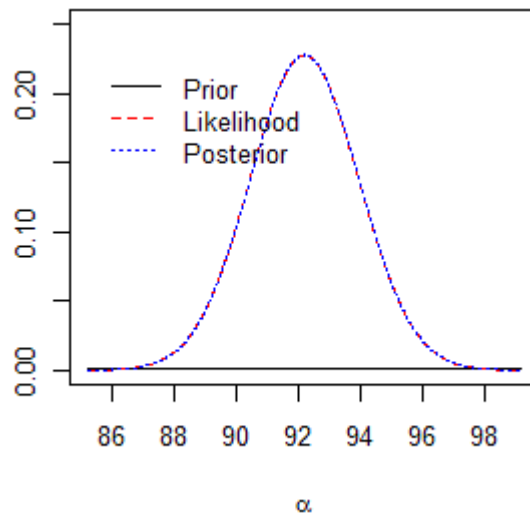
	-----	-----
Intersección:	92.2	1.7535
Pendiente:	1.968	0.2021

Además de las gráficas:

Gráfica 4.2.7



Gráfica 4.2.8



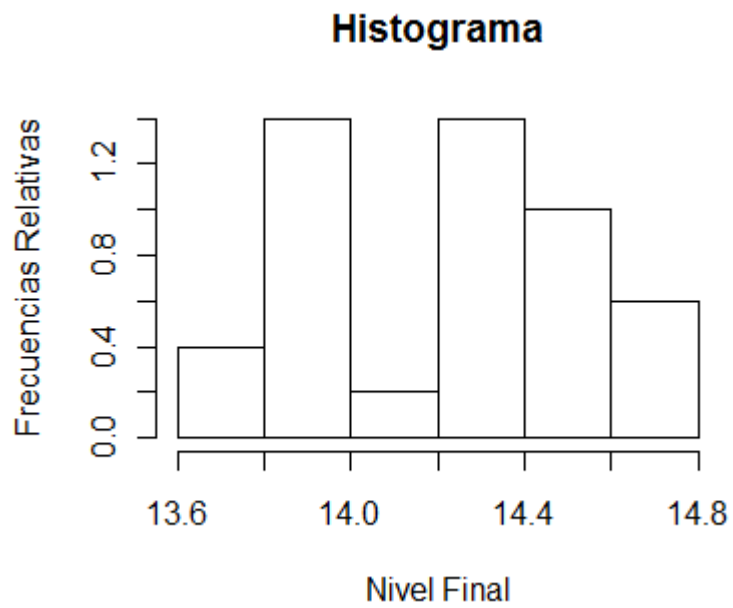
En las cuales no hay diferencia alguna entre las tres curvas normales.

### 4.3 Ejemplo 3

Se utilizan los datos del ejemplo 3 del capítulo 3.

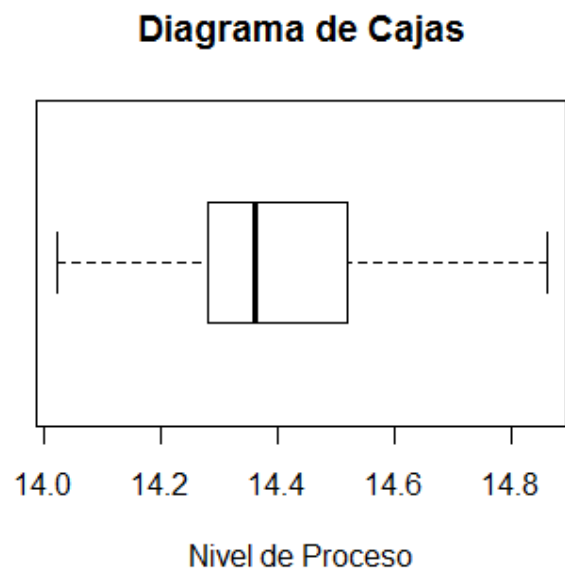
El histograma del nivel final es:

Gráfica 4.3.1



El diagrama de cajas para la variable independiente es el siguiente:

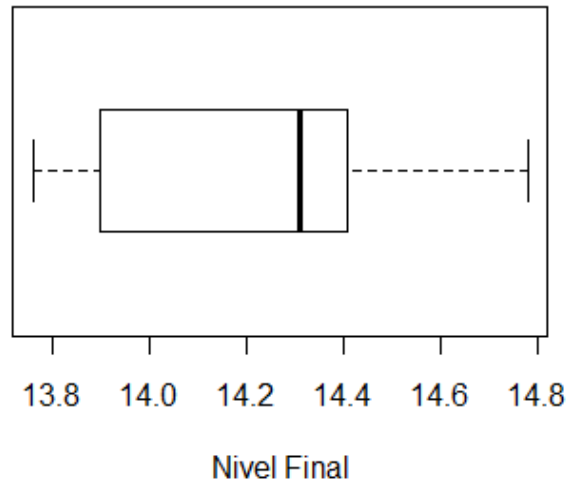
Gráfica 4.3.2



Y el diagrama de cajas para la variable dependiente es:

Gráfica 4.3.3

### Diagrama de Cajas



Ahora se procede a calcular los datos necesarios para obtener la ecuación de regresión:

$$\bar{x} = 14.3888, \bar{y} = 14.2208, \overline{x^2} = 207.0703, \overline{xy} = 204.6628.$$

Luego, se calcula la recta por mínimos cuadrados relativa al nivel final de humedad. La pendiente está dada por:

$$B = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{204.6628 - 14.3888 \times 14.2208}{207.0703 - (14.3888)^2} = \frac{0.0425690}{0.032756} = 1.29963.$$

La recta por mínimos cuadrados es:

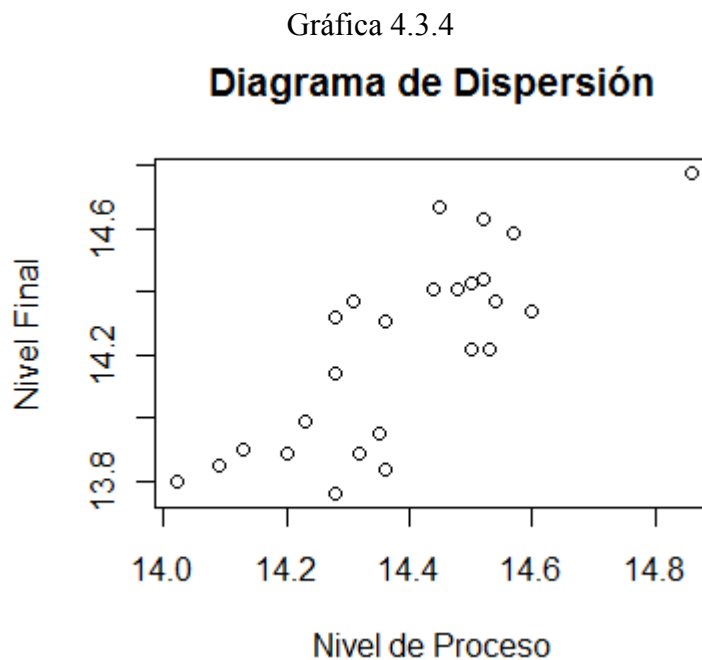
$$y = \bar{y} + B(x - \bar{x}) = 14.2208 + 1.29963 \times (x - 14.3888).$$

Después, se decide que se utilizará una función *a priori normal*(1, (0.3)<sup>2</sup>) para  $\beta_1$  y una *a priori normal*(15, (0.3)<sup>2</sup>) para  $\alpha_x$ .

El programa en para ver la dispersión de los datos es:

```
x<-c(14.36, 14.48, 14.53, 14.52, 14.35, 14.31, 14.44, 14.23, 14.32,14.57, 14.28, 14.36,
14.5, 14.52, 14.28, 14.13, 14.54, 14.6, 14.86,14.28, 14.09, 14.2, 14.5, 14.02, 14.45)
y<-c(13.84, 14.41, 14.22, 14.63, 13.95, 14.37, 14.41, 13.99, 13.89,14.59, 14.32, 14.31,
14.43, 14.44, 14.14, 13.9, 14.37, 14.34, 14.78,13.76, 13.85, 13.89, 14.22, 13.8, 14.67)
plot(x,y,main="Diagrama de Dispersion",xlab="Nivel de Proceso",ylab="Nivel Final")
```

Lo que resulta es el siguiente gráfico:



El coeficiente de correlación entre ambas variables es 0.7956151.

Ahora, al correr el programa:

```
bayes.lin.reg(y, x, slope.prior = "normal", intcpt.prior = "normal", mb0 = 1, sb0 = 0.3,
ma0 = 15, sa0 = 1, sigma = NULL, alpha = 0.05, plot.data = TRUE, pred.x = NULL, ret
= FALSE)
```

Los resultados son:

Tabla 4.3.1

Desviación estándar de los residuos: 0.187

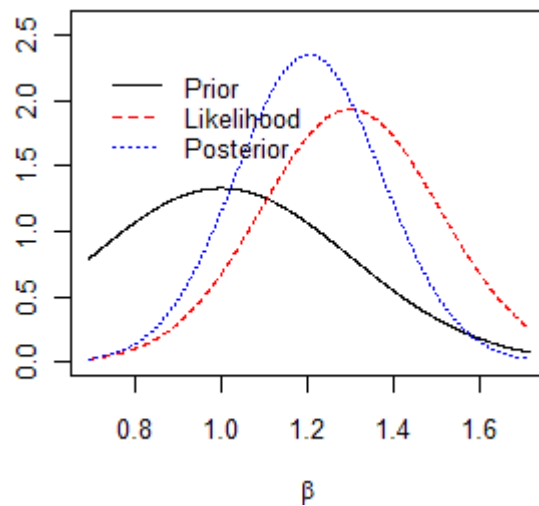
Media Posteriori    Desv. Est. Posteriori

	-----	-----
Intersección:	14.22	0.037318
Pendiente:	1.203	0.17001

Entonces, se tiene una distribución normal *posteriori* con media 14.22 y desviación estándar de 0.037318 para la intersección y una normal *posteriori* con media 1.203 y una desviación estándar igual a 0.17001 para la pendiente. Asimismo se obtuvo una estimación de la desviación estándar de los residuos igual a 0.187.

También las gráficas para la distribución *a priori* y *posteriori* de la pendiente y de la intersección, resultaron de la corrida.

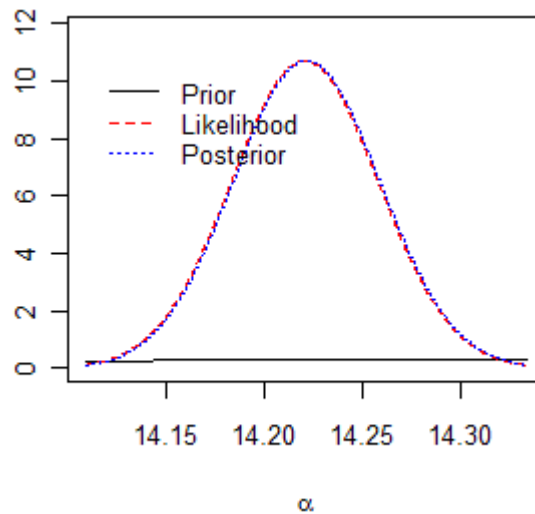
Gráfica 4.3.5



Distribución *a priori*, *posteriori*  
y la *función de verosimilitud* de la pendiente



Gráfica 4.3.6



Distribución *a priori*, *posteriori* y la *función de verosimilitud* de la intersección

Esto dice que no hay diferencias entre las normales de la *función de verosimilitud* y la *posteriori*, además la *a priori* es una constante.

Un intervalo de credibilidad al 95% para la pendiente, es el intervalo (0.852, 1.555). Al igual que en los dos problemas anteriores, el intervalo de credibilidad no toma al cero, y se rechaza la hipótesis nula de que la pendiente sea cero, por lo que se sostiene que el nivel de humedad final puede ser estimado utilizando la medida dentro del proceso de nivel de humedad.

Luego, se hace con la *a priori* no informativa.

**bayes.lin.reg**(y, x, slope.prior = "flat", intcpt.prior = "flat", mb0 = 1, sb0 = 0.3, ma0 = 15, sa0 = 1, sigma = NULL, alpha = 0.05, plot.data = TRUE, pred.x = NULL, ret = FALSE)

Se tienen los resultados:

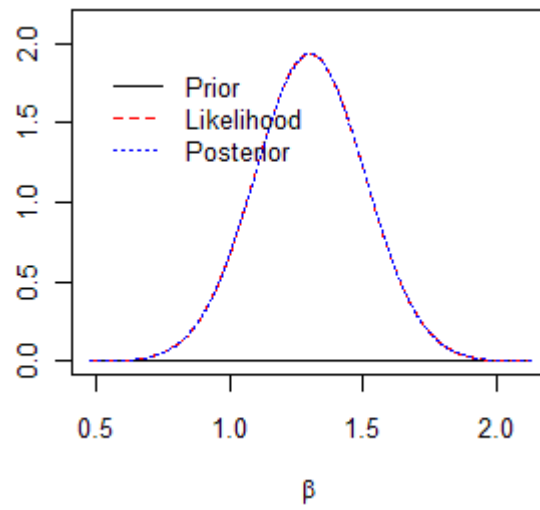
Tabla 4.3.2

Desviación estándar de los residuos: 0.187

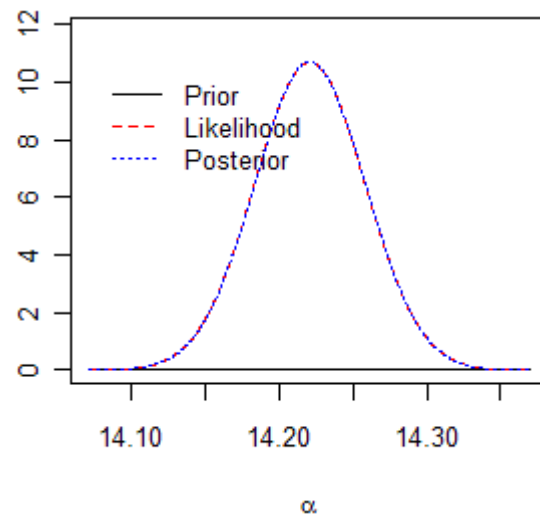
Media Posteriori    Desv. Est. Posteriori

	-----	-----
Intersección:	14.22	0.037344
Pendiente:	1.3	0.20634

Gráfica 4.3.7



Gráfica 4.3.8



No hay diferencias entre las curvas.

## Resultados y conclusiones

De los análisis hechos se desprenden varias consecuencias.

Para el ejemplo 1, en la regresión lineal simple frecuentista los coeficientes para la intersección y la pendiente fueron 26.18623 y -0.40374, respectivamente. Además, con una confianza del 95% el valor de la intersección está dentro del intervalo (23.81, 28.62) y el de la pendiente está dentro de (-0.59, -0.21). El ancho del primer intervalo es de 4.81 y el ancho del segundo es 0.38. El coeficiente de la desviación estándar estimado fue de 4.9752.

En la regresión lineal simple bayesiana para una distribución normal los coeficientes obtenidos fueron 23.01 para la intersección y -0.4021 para la pendiente, además para la desviación estándar estimada se obtuvo uno de 4.98. El intervalo de credibilidad al 95% que se encontró para la pendiente fue (-0.53859, -0.26552), el rango de éste es de 0.8041.

Si se considera una *a priori* no informativa se obtiene un coeficiente para la intersección igual a 23.01 y para la pendiente -0.4037 además de un coeficiente de desviación estándar estimada de 4.98.

En el ejemplo 2, para la regresión lineal simple frecuentista se tiene que los coeficientes de regresión son: -106.133 y 1.968. Un intervalo de confianza que se obtuvo al 95% para la intersección y para la pendiente es (-149.204, -62.607) y (1.538, 2.394), respectivamente. Los rangos de estos son 86.597 y 0.856. El coeficiente de la desviación estándar estimado es 8.7676.

En la regresión bayesiana se obtuvieron los coeficientes 92.21 y 1.972 para la intersección y la pendiente respectivamente y el coeficiente de 8.77 para la desviación estándar estimada. Un intervalo de credibilidad al 95% para la pendiente es (1.68958, 2.27775) y el rango de este intervalo es 0.58817.

Para una función *a priori* no informativa se obtuvo 92.2 y 1.968 que son los coeficientes respectivos de la intersección y la pendiente, y el valor de la desviación estándar estimada es de 8.77.

Por último, para el ejemplo 3 los coeficientes que se obtuvieron para el modelo lineal simple antes de hacer la regresión sin la intersección (pues resultó que no era un parámetro significativo) fueron: -4.47938 y 1.29963 y al sacar los intervalos de confianza al 95% el valor de la intersección está dentro del intervalo (-10.08, 1.51) y el valor de la pendiente dentro de (0.88, 1.68). El rango de estos intervalos son: 8.57 y 0.8 respectivamente. El coeficiente de la desviación estándar estimado es 0.18672.

En la regresión sin la intersección se tiene lo siguiente: el coeficiente de la pendiente fue de 0.988373, resultado que en la prueba de significancia éste es un coeficiente significativo para la regresión. El coeficiente de la desviación estándar fue de 0.19162.

En la regresión bayesiana se obtuvieron los coeficientes 14.22 y 1.203 para la intersección y para la pendiente, un intervalo de credibilidad al 95% para la pendiente es (0.852, 1.555), el rango de este es 0.703. El coeficiente de la desviación estándar estimada es de 0.187.

Para una distribución *a priori* no informativa se tiene un coeficiente de 14.22 y de 1.3 respectivamente para la intersección y para la pendiente, además de un coeficiente de la desviación estándar de 0.187.

## **Conclusiones de los resultados**

Hay que recordar que el objetivo de la tesis no es el optimizar los coeficientes de regresión, sino comparar y comprobar si existe alguna o algunas diferencias entre los métodos de regresión lineal simple frecuentista y bayesiano. Esto se pudo observar al seguir el proceso de cada uno de los métodos teniendo en cuenta los resultados que se obtuvieron.

Los supuestos para ambos métodos son los mismos, pues los dos admiten normalidad de los datos, además de los supuestos de independencia de los errores, etc.

Los coeficientes de la intersección y la pendiente en cada caso son esencialmente los mismos, pues hay que recordar que el estimador de la intersección en el método frecuentista es  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$  y para el bayesiano se utilizó solamente a  $\bar{y}$  para representar la intersección en la ecuación de regresión. Por lo que las ecuaciones de regresión de cada uno de los ejemplos son las mismas. Pero para que sea válida esta conclusión hay que notar que son los mismos coeficientes por el tipo de distribución *a priori* que se utilizó, ya que si se utilizara otra distribución *a priori* seguramente se obtendrían resultados muy distintos. Igualmente el coeficiente estimado de la desviación estándar es el mismo en cada caso.

En cuanto a los intervalos de confianza y los de credibilidad, aunque se interpreten de manera distinta, ya que los primeros miden la confianza de que un cierto valor se encuentre dentro de un rango de valores y los intervalos de credibilidad miden la probabilidad de que un valor esté dentro del mismo, es posible sustentar que los dos deben de contener al verdadero valor del parámetro.

Se puede señalar que los ejemplos realizados apoyan el hecho de que dependiendo del tipo de *a priori* que se utilice, se obtendrán distintos resultados, estos pudieran ser iguales o diferentes a los que en la regresión frecuentista resulten. Por lo tanto, en estos ejemplos no hubo una diferencia entre los métodos, dado que se consiguieron casi los mismos coeficientes para la pendiente, intersección y desviación estándar. Además, se puede decir que ambos métodos tratan de explicar el comportamiento de los datos de la misma forma, esto es a través de una recta de regresión. Pero es posible ver que existe diferencia en cuanto a como se lleva a cabo cada uno de estos procesos de análisis, pues la regresión bayesiana al manejar probabilidades subjetivas tiene un mayor riesgo de que el investigador cometa errores al proponer la distribución *a priori* de los parámetros, pero también tiene ventajas, pues se podría utilizar información previa que en la estadística clásica sería simplemente desechada, ya que ésta utiliza una *a priori* fija en todo el proceso.

En regresión bayesiana, al igual que en la clásica se puede realizar la prueba de correlación entre variables determinando el coeficiente de Pearson, por lo que no hay diferencia entre un método y el otro en esta cuestión.

## **Conclusiones generales**

La estadística que se utiliza comúnmente es la frecuentista, que es la que se desarrolla a partir de los conceptos de probabilidad y que se centra en el cálculo de probabilidades y en la contrastación de hipótesis. De alguna forma, la estadística clásica tiene como objetivo determinar una conclusión, sea en base a significación estadística o aceptación y rechazo de hipótesis, siempre dentro del marco del estudio que se esté realizando. En el análisis estadístico que pretende comparar la eficacia de un nuevo tratamiento frente a otro conocido, se utiliza únicamente la información obtenida de la muestra, que son los datos. No existen subjetividades referentes a parámetros, puesto que se han fijado los criterios de decisión *a priori* y estos permanecen estáticos durante todo el estudio.

En el enfoque alternativo a la estadística frecuentista, aparece cada vez más en escena la estadística bayesiana, basada como su nombre indica en el teorema de Bayes, y que se diferencia de la primera básicamente en la incorporación de información externa al estudio que se esté realizando, de manera que, si se conoce la probabilidad de que ocurra un suceso, su valor será modificado cuando se disponga de esa información. Así pues, las fuentes de información *a priori* se ven transformadas en probabilidad *posteriori* y se utilizan para realizar la inferencia.

La regresión bayesiana es un método muy sensitivo, pues al formar parte de una estadística subjetiva contenida en una ambiente de probabilidades, se podrían presentar varios problemas al momento de hacer un análisis. Ya que al elegir una distribución *a priori* de los coeficientes de regresión se pueden cometer errores al momento de tomar estas decisiones como investigador y llegar a un modelo que realmente no ajuste bien los datos. Por lo que la regresión bayesiana depende de que se tenga una buena información *a priori* para así poseer una distribución *posteriori* efectiva.

En el análisis estadístico, la metodología bayesiana, comienza resumiendo cuantitativamente la información previa existente de origen diverso, desde datos de laboratorios a través de experimentos, estudios publicados e incluso opinión de expertos. Y de esto, se amplía a otro contexto que hace referencia a la definición subjetiva de probabilidad, es decir, a la convicción personal del investigador para emitir juicios sobre una hipótesis. Sin embargo, el método frecuentista no está exento de subjetividad, puesto que el nivel de significancia, así como las explicaciones de la causalidad que se concede a determinados resultados apuntan hacia una apreciación propia de cada investigador.

El área de aplicación de la metodología bayesiana es la misma que la de la estadística clásica, pero ofrece mayores ventajas en distintas situaciones. Simplemente son dos métodos para el análisis estadístico.

En estudios actuales se aplica la metodología frecuentista, aunque en muchas ocasiones, el razonamiento coincide más con el bayesiano. Posiblemente, no sea muy importante elegir un método u otro como herramienta de investigación. Pero el hecho de poder analizar los resultados y entender lo que se está haciendo, en contraposición a una aplicación mecánica del enfoque clásico, dificultan la elección de la metodología bayesiana.

## **Bibliografía**

**Montgomery, D. and Peck, E. A.**(1991): *Introduction to Linear Regression Analysis*. Second edition. Arizona State University

**Ardanuy, A. R.; Etxeberria, J.; Rodríguez, G. G. y Tejedor, T. F.** (2001): *Cuadernos de Estadística*. Editorial: La Muralla S.A. España.

**Ruíz Maya Pérez, L; Martín Pliego F. J.** (1996): *Fundamentos de Inferencia Estadística*. Editorial AC, libros científicos y técnicos de Madrid. Segunda Edición

**Birkes, David; Dodge, Yadolah.** (1993): *Alternative Methods of Regression*. Editorial: Wiley - Interscience.

**Draper, N. R. & Smith, H.** (1998): *Applied Regression Analysis*. Tercera edición, John Wiley and Sons.

**Neter, J., Kutner, M. H., Nachtsheim, C. J. & Wasserman, W.** (1996): *Applied Linear Statistical Models*. Cuarta edición, McGraw-Hill.

**Box, G.E.P. y Tiao, G.C.** (1973): *Bayesian Inference in Statistical Analysis*. Reading: Addison-Wesley.

**Bernardo, J.M. y Smith, A.F.M.** (1994): *Bayesian Theory*. Chichester: Wiley.

**Press, S.J.** (1989): *Bayesian Statistics. Principles, Models and Applications*. New York: Wiley.

**M. Bolstad, William.** (2007): *Introduction to Bayesian Statistics*. Editorial: Wiley - Interscience. Second Edition

**Gelman, A., Carlin, J.B., Stern, H.S. y Rubin, D.B.** (1995): *Bayesian Data Analysis*. London: Chapman & Hall.



**Congdon, P.** (2003): *Applied Bayesian Modelling* Chichester: Wiley.

**Berry, D.A.** (1996): *Statistics: A Bayesian Perspective*. Duxbury Press: Belmont.

**Sivia, D.S.** (1996): *Data Analysis: A Bayesian Tutorial*. Clarendon Press: Oxford.

**Mitchell and Beauchamp** (1988): *Bayesian variable selection in linear regression*. JASA

### **Páginas Web**

<http://www.isds.duke.edu/courses/Summer08/sta103/notes/notes/bayesian.pdf>

<http://www.laflecha.net/articulos/blackhats/estadistica/>

[http://sisbib.unmsm.edu.pe/bibvirtualdata/Tesis/Basic/yupanqui\\_pr/capitulo4.PDF](http://sisbib.unmsm.edu.pe/bibvirtualdata/Tesis/Basic/yupanqui_pr/capitulo4.PDF)

[http://issuu.com/byrong/docs/modelacion\\_regresion](http://issuu.com/byrong/docs/modelacion_regresion)

[http://bochica.udea.edu.co/~bcalderon/3\\_metodosestimacion.html](http://bochica.udea.edu.co/~bcalderon/3_metodosestimacion.html)

<http://personal.us.es/vararey/adatos2/correlacion.pdf>

<http://ecologiadelpaisaje.googlepages.com/Tema15.doc>

<http://www.monografias.com/trabajos27/regresion-simple/regresion-simple.shtml>

[http://www.udc.es/dep/mate/estadistica2/sec6\\_3.html](http://www.udc.es/dep/mate/estadistica2/sec6_3.html)

<http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/Bayes/tema1bayes.pdf>

<http://www.seh-lelha.org/bayes1.htm>

<http://www.mat.ucm.es/~villegas/Porque.pdf>

<http://www.math.mcmaster.ca/canty/teaching/stat743/lectures7.pdf>

[http://lstat.kuleuven.be/java/version2.0/Content\\_Basics.htm](http://lstat.kuleuven.be/java/version2.0/Content_Basics.htm)

<http://ite.gmu.edu/~klaskey/SYST664/SYST664.html>

[http://ite.gmu.edu/~klaskey/SYST664/Bayes\\_Unit4.pdf](http://ite.gmu.edu/~klaskey/SYST664/Bayes_Unit4.pdf)

[http://ite.gmu.edu/~klaskey/SYST664/Bayes\\_Unit8.pdf](http://ite.gmu.edu/~klaskey/SYST664/Bayes_Unit8.pdf)

<http://www.seden.org/files/7-CAP%207.pdf>

<http://predictive.wordpress.com/2008/04/08/no-rotundo-a-la-estadistica-bayesiana/>