

**ALGUNOS EFECTOS DE LOS PUNTOS DE INFLUENCIA  
Y TRANSFORMACIONES LINEALES DE LAS VARIABLES  
PREBICTORAS EN METODOS DE SELECCION DE  
VARIABLES**

**GABRIEL SUAREZ**

**T E S I S**

*Presentada como requisito*

*parcial para obtener el grado de:*

**MAESTRO EN CIENCIAS**

*Especialidad Estadística Experimental*

1984

## CAPITULO 1.

### INTRODUCCION.

Los criterios de selección y métodos computacionales para determinar el mejor subconjunto de variables a ser usada en una ecuación de regresión múltiple, tuvo un gran desarrollo en las pasadas dos décadas. En la actualidad esta área recibe menos atención en la literatura, entre otras cosas, debido a la existencia de muy eficientes métodos de selección de variables.

El problema de determinar el "mejor" subconjunto de variables ha tenido gran aplicación especialmente en estos últimos tiempos debido principalmente a la facilidad que ofrece la existencia de computadoras de alta velocidad, así como también la gran cantidad de paquetes estadísticos disponibles (SAS, SPSS, P-STAT, etc.); lo anterior ha permitido que usuarios con pocos conocimientos estadísticos tengan acceso directo a estos métodos, sin realizar un estudio previo de la estructura de los datos, lo que ha dado como resultado que tales personas no hayan tenido todos los beneficios que proporcionan los métodos de selección de variables.

Hocking (1976), indica que el problema de seleccionar un subconjunto de variables predictoras se encuentran enmarcado bajo dos suposiciones:

- a). El analista tiene datos en un gran número de variables potenciales, las cuales incluyen todas las variables - importantes, algunas funciones apropiadas de ellas y - otras variables no importantes; y,
- b) El analista tiene disponible datos "buenos" sobre los - cuales basarán sus conclusiones. En la práctica si no se satisfacen dichas suposiciones, el analista convertirá un buen proceso de selección de variables en un - ejercicio sin sentido.

El problema de asegurarse que el "conjunto de variables" - contengan todas las variables importantes así, como funcio nes de ellas no es en la práctica un tarea fácil. El ana- lisis de residuales (Anscombe (1961), Draper y Smith (1966); y Daniel y Wood (1971)) puede revelar diferentes formas -- funcionales a ser consideradas y pueden sugerir variables que inicialmente no estaban incluidas, pero este proceso so lo lo puede llevar a cabo un analista de datos experimenta- do.

Es importante indicar que la suposición de "datos buenos" indicados por Hocking, incluye las suposiciones usuales del modelo lineal como son homogeneidad de varianzas, normali- dad e independendencia. El análisis de residuales puede reve lar datos malos, puntos aberrantes o puntos de influencia.

Es bien conocido que una simple observación o conjunto de observaciones, pueden influenciar fuertemente las inferencias y conclusiones que se pueden extraer del análisis de regresión y de esta manera afectar directamente en la selección de variables.

Muchos textos difieren en los métodos para probar la significancia de los coeficientes de menor orden en los modelos de regresión polinomial. Una razón de esta diferencia se debe probablemente que la estadística  $t$  para los coeficientes de regresión de más bajo orden pueden ser cambiados como un resultado de una transformación lineal, Griepentrog et al (1982).

Al modificarse la razón  $t$  para los coeficientes de regresión de más bajo orden los resultados obtenidos usando los datos sin transformar y transformados serán diferentes - cuando se usan métodos de selección de variables (Stepwise).

## 1.2. Objetivo.

El objetivo principal de este trabajo de investigación, es el de presentar un estudio sobre la detección y manejo de algunos factores que afectan a los resultados que se obtienen mediante los métodos más usuales de selección de variables.