



### EXAMEN IN 540

Lunes 11 de Julio de 2005

Profesores: Gonzalo Reyes - Marcelo Henríquez

Auxiliar: José Ignacio Sembler

1. Se dispone de información económica-financiera acerca de un conjunto de empresas del sector manufacturero. Las variables disponibles son:

|         |                              |       |                         |
|---------|------------------------------|-------|-------------------------|
| ACTIVOT | activos totales              | AC_PC | razón de solvencia      |
| AR_PE   | razón de garantía            | FPROP | fondos propios          |
| RE      | rentabilidad económica       | RF    | rentabilidad financiera |
| VENTAS  | ventas totales del ejercicio |       |                         |

Al aplicar un Análisis Factorial, la matriz de componentes rotada una vez extraídos 3 factores es:

|         | Factor 1 | Factor 2 | Factor 3 |
|---------|----------|----------|----------|
| ACTIVOT | 0.982    | -0.085   | 0.068    |
| AC_PC   | -0.306   | 0.826    | 0.154    |
| AR_PE   | 0.110    | 0.918    | 0.170    |
| FPROP   | 0.934    | 0.157    | 0.041    |
| RE      | -0.053   | 0.072    | 0.981    |
| RF      | 0.347    | 0.378    | 0.835    |
| VENTAS  | 0.808    | -0.275   | 0.113    |

- Haga una interpretación general del resultado, tanto de la bondad del modelo como de los factores (la ordenación que producen en las empresas). (10 puntos)
- ¿Cómo es una empresa cuyas puntuaciones factoriales son (2.823,-1.199, 0.577)? (10 puntos)
- Si el investigador decide comparar esta estructura factorial con la de otros sectores económicos, ¿le sugiere usted estandarizar las variables antes de efectuar el AF? (10 puntos)

#### **Solución P.1.**

- La media de las 7 comunalidades nos da una medida de la proporción de varianza explicada por el modelo, en este caso es 0.89, buen resultado para 7 variables.  
Primer eje: ordenará las empresas principalmente en función de las variables activo total, fondos propios y ventas, de mayor valor (cuanto más activo, fondos propios y ventas tenga) a menor (lo opuesto).  
Segundo eje: ordenará las empresas principalmente en función de las razones de solvencia y garantía, de mayor (cuanto mayores sean esas razones) a menor (lo opuesto).  
Tercer eje: ordenará las empresas principalmente en función de las variables de rentabilidad, de mayor (cuanto mayores sean dos variables) a menor (lo opuesto).*
- La empresa con las puntuaciones mencionadas es de enorme envergadura, de baja solvencia y poco destacada en rentabilidad (pero por encima de la media)*
- Sí. Aunque los resultados de un AF no son afectados por estandarización, la cual se realiza dentro del procedimiento, hay que observar, sin embargo, que la estandarización (sustraer la media, dividiendo por la desviación estándar) escala los datos de una manera específica o particular a la muestra. Si el investigador se*

propone comparar la estructura factorial de dos o más muestras, entonces debería usar la matriz de covarianzas más que la matriz de correlaciones como input del AF. Pero el método en base a la covarianza tiene problemas cuando las variables son muy diferentes en escala (ej. Ingreso = \$100,000 y Educación = 21 años). Así, se podría recomendar una estandarización multimuestra (sustraer la gran media de todas las muestras, dividiendo por la desviación estándar de todos los casos) antes de calcular la matriz de covarianza de las muestras. Con todo, en la práctica la comparación de factores entre muestras está basado en el típico AF basado en matrices de correlación.

2. Un banco aplica Análisis Discriminante sobre una muestra representativa de clientes a los cuáles prestó dinero, para estudiar cuál de sus características Patrimonio Neto o Deuda Pendiente, es más determinante a la hora de observar el resultado del préstamo (dicotomizado en Fallido o Devuelto). Los siguientes son los principales resultados del AD estimado con SPSS:

Wilks' Lambda

| Test of Function(s) | Wilks' Lambda | Chi-square | df | Sig. |
|---------------------|---------------|------------|----|------|
| 1                   | ,368          | 12,987     | 2  | ,002 |

Eigenvalues

| Function | Eigenvalue         | % of Variance | Cumulative % | Canonical Correlation |
|----------|--------------------|---------------|--------------|-----------------------|
| 1        | 1,716 <sup>a</sup> | 100,0         | 100,0        | ,795                  |

a. First 1 canonical discriminant functions were used in the analysis.

Standardized Canonical Discriminant Function Coefficients

|                 | Function |
|-----------------|----------|
|                 | 1        |
| Patrimonio neto | ,922     |
| Deuda pendiente | -,686    |

Structure Matrix

|                 | Function |
|-----------------|----------|
|                 | 1        |
| Patrimonio neto | ,748     |
| Deuda pendiente | -,452    |

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions  
Variables ordered by absolute size of correlation within function.

Classification Function Coefficients

|                 | Devolución del préstamo |          |
|-----------------|-------------------------|----------|
|                 | Fallido                 | Devuelto |
| Patrimonio neto | ,777                    | 1,813    |
| Deuda pendiente | 1,296                   | ,364     |
| (Constant)      | -5,876                  | -9,396   |

Fisher's linear discriminant functions

- ¿Cuál característica es más determinante en la probabilidad de devolución del préstamo? (10 puntos)
- Dados un solicitante A con Patrimonio Neto = 10.1 (MM\$) y Deuda Pendiente = 6.8 (MM\$), y un Solicitante B con Patrimonio Neto = 9.7 (MM\$) y Deuda Pendiente = 2.2 (MM\$), ¿otorgaría o no otorgaría un préstamo a los solicitantes? (10 puntos)
- Para el banco, ¿es lo mismo aplicar un Análisis de Cluster que un AD en este caso? (10 puntos)

## Solución P.2.

- En el cuadro de coeficientes estandarizados de la función discriminante canónica, se observa que el coeficiente de la variable Patrimonio Neto es positivo y el mayor de las dos variables, por lo que un mayor valor de esta variable tendrá una influencia importante en que el préstamo sea devuelto. El coeficiente de la variable Deuda Pendiente es negativo y menor que el de la variable Patrimonio Neto, por lo que una mayor cantidad de prestamos activos influirá negativamente en la posibilidad de devolución del préstamo. Además, en el cuadro de la matriz de estructura, se observa que los signos se mantienen y que la variable Patrimonio Neto comparte una mayor varianza con la función discriminante (0,748) que la variable Devolución Pendiente (-0,452). Esto confirma la interpretación anterior en el sentido que el Patrimonio Neto es más determinante en la probabilidad de devolución de un préstamo.
- Las funciones de Fisher para cada grupo son  
 $F_1 (\text{fallido}) = -5,876 + 0,777 \text{ PN} + 1,296 \text{ DP}$   
 $F_2 (\text{devuelto}) = -9,396 + 1,813 \text{ PN} + 0,364 \text{ DP}$

Y las probabilidades a posteriori de pertenecer a un grupo

$Pr(1/D) = e(F_1) / [e(F_1) + e(F_2)]$  para los fallidos

$Pr(2/D) = e(F_2) / [e(F_1) + e(F_2)]$  para los devueltos

Aplicando a los (nuevos) solicitantes se tiene:

|               | PN (MM\$) | DP (MM\$) | Funciones de Fisher       |                            | Probabilidad a posteriori |                      |
|---------------|-----------|-----------|---------------------------|----------------------------|---------------------------|----------------------|
|               |           |           | F <sub>1</sub><br>Fallido | F <sub>2</sub><br>Devuelto | Pr (1/D)<br>Fallido       | Pr (2/D)<br>Devuelto |
| Solicitante A | 10.1      | 6.8       | 10.8                      | 11.4                       | 35,3%                     | 64.7%                |
| Solicitante B | 9.7       | 2.2       | 4.5                       | 9.0                        | 1.1%                      | 98.9%                |

El resultado de la aplicación de las funciones discriminantes indica que el solicitante A tiene una probabilidad del 35,3% de fallar en la devolución del préstamo, lo que dependiendo de la política de riesgo de la entidad podría ser considerado como un nivel de riesgo demasiado alto para otorgarlo.

En cambio el solicitante B tiene una alta probabilidad de devolver el préstamo y una muy pequeña (1,1%) de no hacerlo, lo que bien podría ser un riesgo aceptable y decidir a favor del otorgamiento por parte del director.

Se aprecia que la diferencia entre los solicitantes está dada por el mayor nivel de endeudamiento del A, ya que ambos tienen niveles de patrimonio similares y que en principio podrían ser suficientes por sí mismos para acceder al préstamo.

- c) No. En el AD los grupos están determinados de antemano y el objetivo es obtener la combinación lineal de las variables independientes que discriminan mejor entre los grupos. En cluster analysis los grupos no están predeterminados y su objetivo es obtener la mejor manera en que los casos se pueden agruparse homogéneamente en base a las variables.

3. Suponga que usted cuenta con el siguiente modelo  $Y_i = \beta_1 + \beta_2 X_{1i} + u_i$

Donde  $Y_i$  puede tomar solamente los valores 0 (cero) o 1 (uno).

- Señale las consecuencias de estimar este modelo de acuerdo al Modelo de Probabilidad Lineal. (10 puntos)
- Plantee el modelo de forma tal que pueda ser estimado mediante el método Probit. Especifique la función de distribución que se utiliza en este caso (No es necesario que entregue la fórmula de esta distribución). (10 puntos)
- Suponga que el modelo planteado anteriormente relaciona la probabilidad de estar ocupado con los años de educación de la persona, de manera tal que: (10 puntos)

$Y_i = 1$  si la persona  $i$  se encuentra ocupado y 0 si no

$X_i$  = años de escolaridad de la persona  $i$

Suponga que se ha estimado el modelo con el método Logit y se encontraron los siguientes resultados:

$$\hat{Y}_i = -0.762787 + 0.0812515 * X_i$$

Error Std= (0,0108358) (0,001114)

- i) Obtenga una expresión para la probabilidad de estar ocupado a partir del modelo estimado
- ii) ¿Cuál es la probabilidad de que una persona con 8 años de escolaridad se encuentre empleada?  
¿Y para alguien con 12 años de educación?
- iii) Obtenga el Efecto Marginal de la escolaridad en la probabilidad de estar empleado para el nivel de educación de 10 años.
- d) Suponga ahora que usted no solo cuenta con información acerca del estatus ocupacional de la persona sino que además observa el número de horas trabajadas para aquellos que se encuentran ocupados. De este manera la nueva variable dependiente toma los siguientes valores:  
 $Y_{1i}$  = horas trabajadas si la persona está ocupada y  
 0 si la persona no está ocupada.  
 Explique cómo estimaría este modelo como un modelo tobit siguiendo la metodología de dos etapas de Heckman. (10 puntos).

### **Solución P.3.**

- a) Hay cuatro consecuencias de estimar de acuerdo al modelo de probabilidad lineal: los errores no son normales, son heteroscedásticos, el  $R^2$  no es una buena medida de bondad de ajuste y no se cumple necesariamente que la  $E(Y/X)$  este entre 0 y 1.

En efecto,  $Y_i$  sólo puede tomar valores 0 o 1 (éste último con probabilidad  $P_i$ ), entonces

$$u_i = \begin{cases} -(\beta_1 + \beta_2 X_{1i}) & \text{con probabilidad } 1 - P_i \\ 1 - (\beta_1 + \beta_2 X_{1i}) & \text{con probabilidad } P_i \end{cases}$$

O sea,  $u_i$  no es v.a. normal sino Bernoulli.

Recordando que la varianza de una v.a. con esa distribución es  $P_i(1 - P_i)$ , la cual depende de  $i$ , es decir no es fija, entonces hay heteroscedasticidad.

El problema con  $R^2$  como medida de bondad de ajuste puede verse gráficamente. Como  $Y_i$  sólo puede tomar valores 0 o 1, la suma de los cuadrados residuales no es muy informativa. El ajuste de la línea será bueno para los puntos cercanos a esos valores y malo para puntos intermedios.

Por último, como estimamos una función lineal con pendiente distinta de cero, ésta seguramente tomará valores inferiores a cero y superiores a 1 para algún rango de valores de  $X_{1i}$ .

4. Suponga que usted tiene las siguientes funciones de Oferta y Demanda:

Función de Demanda:  $Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t}$

Función de Oferta:  $Q_t = \beta_0 + \beta_1 P_t + u_{2t}$

- a) Encuentre si estas ecuaciones están sub, sobre o exactamente identificadas. (10 puntos)
- b) ¿Qué condición debe cumplirse para que exista identificación de la ecuación estructural a partir del sistema de ecuaciones? Muestre como esta condición se cumple o incumple en el caso anterior. (10 puntos)

**TIEMPO: 120 Minutos**