

FACTORES IMPORTANTES EN EL EXAMEN DE ADMISIÓN-UNI¹

Alipio Ordoñez Mercado²

Facultad de Ingeniería Económica y Ciencias Sociales

Universidad Nacional de Ingeniería Lima - Perú

RESUMEN

En este trabajo se investiga las variables explicativas más relevantes que describen, al comportamiento del puntaje en el examen de admisión a la Universidad Nacional de Ingeniería. Para ello se usaron dos técnicas estadísticas; la "Regresión Stepwise", y la "Regresión Ridge". En el análisis se utilizó los datos correspondientes al examen de admisión para el año de 1994. La recolección de los datos se obtuvo mediante un muestreo estratificado sistemático con asignación proporcional al tamaño poblacional. Dicho análisis revela que la variable más importante recae en el "PROMEDIO GENERAL" (X_6), obtenido durante la educación secundaria; entre otras de las variables más relevantes se encuentran: El "Canal de Ingreso" (X_1), el "Tipo de preparación Pre-Universitaria" (X_4), y el "Tiempo de preparación Pre-Universitaria" (X_7).

ABSTRACT

In this paper the most important explicative variables in the behaviour of admission exam grades to National Engineering University is investigated. The techniques STEPWISE REGRESSION and RIDGE REGRESSION are considered and in the analysis is used the results of 1994 admission exam. These data were obtained through sistematic stratificated sampling with proporcional asignation to size population. The results reveals that the most important variable is the general average in high school (X_6); others important variables are Admission canal (X_1), Pre-University Preparation type (X_4), and Pre-University Preparation Time (X_7).

INTRODUCCIÓN

Muchos de los trabajos de investigación estadística en el campo educacional, han sido direccionadas a la forma como se debe de medir el rendimiento académico de los estudiantes en las diversas instituciones dedicadas al proceso de la enseñanza en sus niveles diferentes. En particular la evaluación del rendimiento académico de los postulantes al examen de admisión en la Universidad Nacional de Ingeniería, es un tema interesante que merece un estudio exhaustivo y detallado en todas sus implicancias. Entre muchos de los posibles estudios, la identificación de las variables relevantes que tienen una marcada influencia sobre el "PUNTAJE FINAL" con el que ingresan a la universidad; es de especial interés para quienes tienen bajo su responsabilidad la decisión de captar los mejores estudiantes para la Universidad.

Entre el grupo de variables que tienen cierta influencia en los puntajes de admisión, se perfilan las

notas o calificaciones obtenidas por el candidato en los colegios de Educación Secundaria. Es de conocimiento de muchas personas, que un examen de admisión en la UNI es relativamente difícil en comparación a otros exámenes similares; razón por el cual puede suponerse que los estudiantes en su mayoría tienen una base educacional en la secundaria y/o pre-universitaria muy buena, esto prácticamente estaría sugiriendo incluirlas como indicadores en la decisión final de todo proceso de admisión a la Universidad.

Este estudio intentará especificar e identificar a los factores más relevantes en la obtención de un buen puntaje en el examen de admisión, para ello se buscará variables que estén fuertemente asociados a los puntajes del examen de admisión; Asimismo se explotará esta asociación existente para formular un modelo que los relaciona. El método tradicional para realizar la selección de las principales variables, "REGRESIÓN STEPWISE", y La "REGRESIÓN RIDGE", para los casos en que se tiene

¹ Trabajo de investigación subvencionado por la Oficina central de admisión-UNI. Nov. 1994.

² Prof. Asociado de Escuela Profesional de Estadística en la Facultad de Ingeniería Económica y Ciencias Sociales- Univ. Nac. de Ingeniería. Lima- PERÚ.

multicolinealidad se usarán e implementarán para realizar el análisis de los datos correspondientes.

FACTORES INFLUYENTES EN EL EXAMEN DE ADMISIÓN

El rendimiento Académico de todo postulante que pretende ingresar a la Universidad Nacional de Ingeniería puede en términos generales estar afectado, por los siguientes grupos de variables explicativas:

1.- Condiciones de la Educación Secundaria

En este punto se considerarán los antecedentes que el postulante trae, desde el escenario de la educación secundaria por ejemplo:

- Calificaciones en los cursos de: Matemáticas, Física, Química, Castellano, y etc.
- Tipo de colegio,
- Procedencia geográfica,
- Aspecto vocacional y otros más.

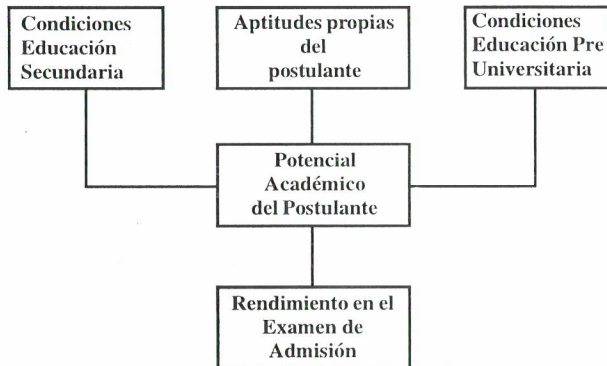


Fig.1: Grupo de Variables que intervienen en Examen de Admisión.

2.- Aptitudes Propias del Alumno

Expresadas en las capacidades propias que el postulante tiene para enfrentar las exigencias del examen de admisión y que los adquiere durante su vida intelectual tanto en el colegio como en el ambiente familiar. Este grupo de variables resultan algo difíciles de ser cuantificados.

3.- Condiciones Pre-Universitarias

Este grupo de variables deben ser cuantificados desde el momento en que el postulante termina sus estu-

dios secundarios hasta el instante en que se lleva a cabo el examen de admisión, por ejemplo:

- Tipo de preparación en la educación pre-universitaria,
- Tiempo de preparación,
- Calificaciones obtenidas en los cursos pre-universitarios,
- Tiempo transcurrido desde que termino la secundaria.

Estos grupos de variables puede ser esquematizados como lo indica la Fig.1. Hojeando rápidamente el expediente de cualquier postulante al examen de admisión; se observará que las posibles variables explicativas bajo la estructura descrita en la Fig.1, y que pueden tener un grado de influencia sobre el puntaje del examen de admisión, se identifican como:

Y : Puntaje de ingreso en el examen de admisión definida en el intervalo $0 \leq Y \leq 20$,

X_1 : Canal (grupos de Especialidades afines) a los cuales se presenta el postulante, y es codificado con los siguientes valores:

3: Canal I; 1: Canal II,
2: Canal III; 4: Canal IV; 5: Canal V,

X_2 : Edad del postulante calculado hasta el mes en que rindió el examen (años),

X_3 : Sexo del postulante, definida por:
1 = Hombre 2 = Mujer,

X_4 : Tipo de preparación pre-universitaria,
1 = Autopreparación,
2 = Prof. Particular,
3 = Academia.

X_5 : Número de veces que postuló a un examen de admisión,

X_6 : Tipo de colegio del cual procede el alumno;
1= estatal, 2= Particular,

X_7 : Tiempo de preparación pre-universitaria (en años);

X_8 : Tiempo transcurrido desde que culminó sus estudios secundarios (años),

X_9 : Procedencia Geográfica,
1= Si vive en Lima Centro,
2 = No vive en Lima centro.

X_{10} : Calificación en Literatura,

X_{11} : Calificación en Aritmética (matemáticas del 1° y 2°),

X_{12} : Calificación en Algebra (matemática 3°),

X_{13} : Calificación en Geometría (matemática 4°),
 X_{14} : Calificación en Trigonometría (matemática 5°),
 X_{15} : Calificación en Química, y
 X_{16} : Calificación en Física.
 Con este grupo de variables se propone el siguiente modelo estadístico:

$$Y = f(X_1, X_2, \dots, X_{16}) + e_i; i = 1, 2, \dots, n.$$

$$Y_i = \beta_0 + \beta_1 X_1 + \dots + \beta_{16} X_{16} + e_i; i = 1, 2, \dots, n$$

Donde los coeficientes beta se interpretan:

β_0 : Efecto de la media general;

$\beta_1, \dots, \beta_{16}$: Efecto de cada una de las variables consideradas.

SELECCIÓN DE MUESTRAS

Los fundamentos de la inferencia estadística se halla fuertemente sustentada, en la representatividad y aleatoriedad de sus elementos que aportan información sobre la característica de interés en la pobla-

ción; en este caso "El rendimiento académico de los postulantes en el examen de admisión", y por esta razón se trató de seleccionar muestras que mejor representan a la población investigada. Para este estudio la población a considerarse es el total de postulantes al examen de admisión en 1994-1, realizado en el mes de Marzo. Escogiéndose un tamaño muestral proporcional al tamaño poblacional en cada canal, y aplicado para las categorías de: ingresantes por primeros puestos, ingresantes ordinarios, y no ingresantes. También se estableció los tercios superior, intermedio e inferior, dentro de la estructura considerada. En esta forma se aplicó un muestreo estratificado Sistemático para cubrir la población en forma uniforme. Las cantidades a ser consideradas se presentan: en la tabla No.1 para los ingresantes por primeros puestos, ingresantes ordinarios, y no ingresantes; en todas ellas se considera los canales y los tercios respectivos. Puede observarse que debido al tamaño moderado de los ingresantes por primeros puestos se decidió considerar el 52%; mientras que en otros tercios se consideró alrededor del 15%, y 5% del tamaño poblacional.

TABLA N°1: Determinación del tamaño muestral - Por Categorías de Ingreso. Examen 1994 - 1

Canal	Población (N_{hr})			Int. Aleatoria(k/a^*)			Peso (W_n)			Puntaje (Max/Min)			Muestra (n_i)			Orden Mérito		
	PP	IO	NI	PP	IO	NI	PP	IO	NI	PP	IO	NI	PP	IO	NI	PP	IO	NI
I	14	83	399	k= 3	7	19	0.163	0.164	0.078	8.075	S= 2	4	7	S= 2-4	1-27	84-217		
											M= 2	4	7	M= 5-9	28-55	218-349		
II	30	170	2004	k= 2	6	20	0.349	0.335	0.389	12.525	S= 5	9	33	S= 1-10	1-56	171-839		
											M= 5	9	33	M=11-20	57-113	840-1506		
III	14	87	1594	k= 3	7	19	0.163	0.172	0.310	11.950	S= 2	4	26	S= 1-4	1-29	88-619		
											M= 2	4	26	M= 5-9	30-58	620-1150		
IV	23	133	1068	k= 2	6	18	0.267	0.262	0.208	9.825	S= 5	7	18	S= 1-8	1-44	134-490		
											M= 5	7	18	M= 9-16	45-89	491-845		
V	5	34	80	k= 2	5	13	0.058	0.067	0.015	5.025	S= 1	2	2	S= 1-2	1-11	35-62		
											M= 1	2	2	M= 3-4	12-23	63-88		
TOTAL	86	507	5145				1.000				45	78	258	52%N	15%N	5%N		

*k: Intervalo sistemático aa: Arranque aleatorio PP: Primeros Puestos IO: Ingresantes Ordinarios NI: No Ingresantes

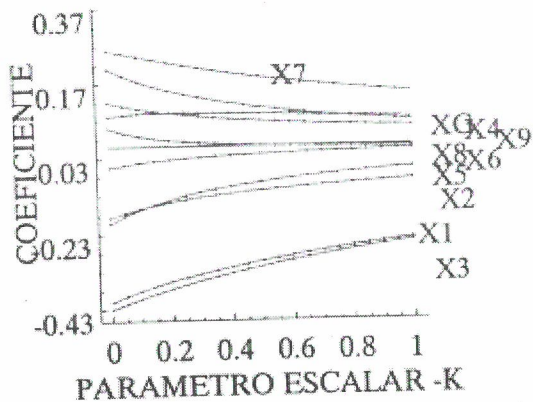


Fig.3: Señal Ridge. Datos: Primeros Puestos. 1994-1

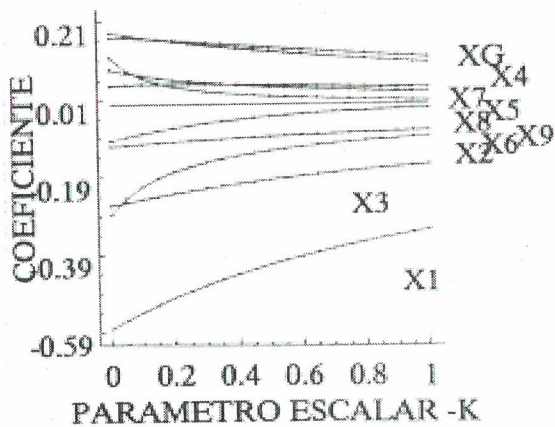


Figura.4: Señal Ridge: Datos Ingresantes ordinarios. 1994-1

significativamente diferente de "0". La inspección gráfica conduce a los resultados presentados en la tabla No. 4.

Los subconjuntos seleccionados, si bien es cierto que son más estables aún conservan su estructura de correlación por lo que algunas deben ser descartadas porque solo bastaría considerar uno de ellos en el modelo. Luego aplicando el criterio de que no exista correlación entre las variables independientes se obtiene las siguientes variables eliminadas.

1.- En los ingresantes por primeros puestos se observa que X_G , está correlacionada con las variables X_2 , X_5 , X_7 , X_8 , y X_9 (Tabla No.5), luego estas serán descartadas, pues X_G representa sus efectos.

También X_1 está correlacionada con X_3 y X_4 , por lo que pueden descartarse, luego en el modelo quedan solo dos variables; X_1 y X_G , como las más importantes y ellas corresponden; al canal que pos-

tulan y al promedio general del colegio del cual proceden.

2.- Un análisis similar en los ingresantes ordinarios conduce a las siguientes observaciones en la

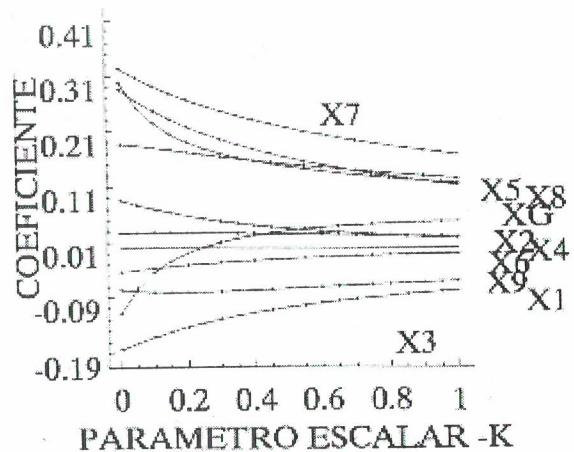


Figura 5: Señal Ridge: Datos No Ingresantes 1994-1.

TABLA No.4:
Variables Seleccionadas vía Regresión Ridge
Datos Examen de Admisión 1994-1

	PRIMEROS PUESTOS	INGRESANTES ORDINARIOS	NO INGRESANTES
VARIABLES EN EL MODELO ESCOGIDO	X_1 *	X_1 *	X_1 *
	X_2 *	X_2 *	X_3 *
	X_3 *	X_3 *	X_5 *
	X_4 *	X_4 *	X_7 *
	X_5 *	X_6 *	X_8 *
	X_7 ns	X_7 *	X_G *
	X_G *	X_G *	
R^2	94.81	97.55	77.29

(*): Significativa al 95%, (ns): No significativa.

Tabla No.6, la variable X_G está correlacionada con las variables $X_1, X_2, X_3, X_5, X_6, X_7, X_8,$ y X_9 ; luego estas serán eliminadas del modelo, y solo permanecerán las variables X_G y X_4 .

- 3.- Aplicando los mismos criterios en los no ingresantes (Tabla No.7), se observará que la variable X_G está correlacionada con las variables $X_2, X_3, X_5, X_7,$ y X_8 ; por que serán descartadas del modelo, también existen correlaciones entre X_1 y X_3, X_6 y X_2, X_3 y X_9 , las cuales fueron eliminadas; así solo quedarían dentro del modelo las variables X_G y X_4 .

Por tanto el Ajuste de estas variables seleccionadas en un modelo de regresión lineal sin el termino constante; proporciona los resultados que se presentan en la tabla No.8. Obsérvese que la variable X_G que identifica el efecto de las calificaciones obtenidas por el postulante en el colegio, y es la más significativa; luego si se incrementa el nivel de confianza, resultará como única variable más importante con un efecto marcado sobre el puntaje general del examen de admisión.

TABLA N.5: Matriz de correlaciones Datos: Ingresantes por Primeros Puestos. Examen 1994-1.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_G
Y	-329	-.245	-.638	.030	-.057	-.000	.318	-.034	.120	.200
X_1		.128	.119	.106	-.304	-.140	-.226	-.093	-.161	.058
X_2			.043	.260	.278	-.149	.245	.490	-.037	-.171
X_3				.085	-.135	-.070	-.183	-.087	-.002	.028
X_4					-.143	.090	.048	-.107	-.234	.005
X_5						-.144	.335	.463	.195	-.243
X_6							-.245	-.161	-.154	-.125
X_7								.505	.009	-.133
X_8									.133	-.160
X_9										.215

(*) : $X_G = (X_{10} + X_{11} + \dots + X_{16}) / 7$

TABLA N.6: Matriz de correlaciones Datos: Ingresantes Ordinarios. Examen 1994-1

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X _G
Y	-0.763	-0.138	-0.391	0.266	0.079	-0.184	0.146	0.023	0.094	0.317
X ₁		0.042	0.285	-0.158	-0.026	0.093	-0.103	-0.063	-0.105	-0.315
X ₂			-0.135	0.107	0.403	-0.267	0.466	0.825	-0.070	-0.424
X ₃				0.080	-0.278	0.114	-0.280	-0.133	-0.171	0.199
X ₄					0.087	-0.077	0.123	0.143	0.089	0.095
X ₅						-0.172	0.377	0.376	0.225	-0.338
X ₆							-0.434	-0.327	-0.022	-0.112
X ₇								0.490	0.085	-0.312
X ₈									-0.137	-0.357
X ₉										0.218

(*) : $X_G = (X_{10} + X_{11} + \dots + X_{16}) / 7$

TABLA N.7: Matriz de correlaciones Datos: No Ingresantes. Examen 1994-1

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X _G
Y	-0.144	0.305	-0.164	0.031	0.410	-0.062	0.451	0.397	0.009	0.142
X ₁		0.100	0.198	-0.036	0.004	-0.049	-0.014	0.046	-0.080	-0.080
X ₂			-0.007	-0.116	0.578	-0.227	0.443	0.820	0.003	-0.158
X ₃				0.072	-0.087	0.071	0.024	-0.050	-0.149	0.180
X ₄					-0.093	-0.009	0.135	-0.103	0.062	0.015
X ₅						-0.190	0.466	0.624	0.076	-0.146
X ₆							-0.171	-0.219	-0.140	-0.053
X ₇								0.468	0.024	-0.108
X ₈									-0.008	-0.207
X ₉										0.040

(*) : $X_G = (X_{10} + X_{11} + \dots + X_{16}) / 7$

TABLA No.8: Modelos Seleccionados: Datos Examen de Admisión 1994-1.

	PRIMEROS PUESTOS	INGRESANTES ORDINARIOS	NO INGRESANTES
VARIABLES EN EL MODELO	$\beta_1 = -1.143$ (5.284)	$\beta_4 = 1.398$ (3.023)	$\beta_6 = 0.129$ (0.617)
	$\beta_6 = 0.557$ (14.769)	$\beta_6 = 0.342$ (3.696)	$\beta_6 = 0.2124$ (4.994)
R^2	91.81	94.96	69.44

Finalmente el ajuste de la variable X_G en el modelo de regresión, explica en más del 70% de variación en los puntajes del examen de admisión a la Universidad Nacional de Ingeniería. Los efectos de las calificaciones del Colegio, también deben verificarse a nivel de los tercios considerados, y estos resultados se confirman en las estimaciones que se presentan en la tabla No.11. Las dos últimas tablas, revelan que la variable X_G , promedio general de las calificaciones obtenidas en los colegios, influyen marcadamente en los puntajes del Examen de Admisión.

TABLA No.9: Las Cinco Variables que más explican individualmente al puntaje del examen. 1994-1

NIVEL GENERAL		PRIMEROS PUESTOS		INGRESANTES ORDINARIOS		NO INGRESANTES	
VAR.	R^2	VAR.	R^2	VAR.	R^2	VAR.	R^2
X_G	71.04	X_G	86.56	X_G	94.41	X_2	71.03
X_8	69.98	X_4	85.36	X_4	94.10	X_G	69.51
X_2	69.27	X_2	85.27	X_2	92.94	X_7	69.23
X_4	68.24	X_9	79.15	X_9	86.83	X_4	66.57
X_9	61.70	X_7	78.50	X_7	83.20	X_9	62.05

TABLA No.10: Efectos de la variable Promedio General, sobre los Puntajes del Examen de Admisión 1994-1.
(Modelo sin término constante.)

	PRIMEROS PUESTOS	INGRESANTES ORDINARIOS	NO INGRESANTES	NIVEL GENERAL
β_G	0.382213	0.618258	0.237985	0.3385
T_{cal}	16.4526	35.3392	24.1117	30.2519
R^2	86.57	94.41	69.51	71.04

TABLA No.11: Efectos de la variable Promedio General, sobre los Puntajes del Examen de Admisión.

1994-1. (Modelo sin término constante.)

CONDICIÓN CANDIDATO	TERCIOS	COEFICIENTE β_G	ESTADÍSTICO $T_{cal} - STUDENT$	R^2 (%)
INGRESANTES ORDINARIOS	SUPERIOR	0.712783	29.1455	97.36
	INTERMEDIO	0.583647	18.5055	93.20
	INFERIOR	0.553162	22.4440	95.45
NO INGRESANTES	SUPERIOR	0.415233	35.2449	93.60
	INTERMEDIO	0.199149	29.0786	91.06
	INFERIOR	0.085588	21.4200	84.37

CONCLUSIONES Y RECOMENDACIONES

El análisis estadístico realizado en la sección 4, revela las siguientes conclusiones y recomendaciones:

- Las calificaciones en los cursos de matemáticas, física, química, castellano (literatura); se encuentran altamente correlacionadas esto sugiere que bastaría con considerar solo una de ellas, pues también explica el efecto de las demás. Luego es más conveniente tomar un promedio general sobre todas las calificaciones obtenidas en el colegio.
- La explicación del promedio general X_G , en la variación de los puntajes del examen de admisión, se ha identificado como la más importante, y tiene una explicación que fluctúa entre el 80% y 97%; Siendo su coeficiente de correlación positiva; lo que indica que a mayor promedio general, mayor será el puntaje a ser obtenido en el examen de admisión.
- El promedio general X_G , obtenido en el colegio está correlacionado con los puntajes del examen de admisión, en porcentajes que varían entre el 14.2 y 31.7, cuando se considera la condición de los postulantes en las categorías de primeros puestos, ingresantes y no ingresantes; encontrando su máxima correlación en los ingresantes y la mínima en los no ingresantes. Obteniéndose un valor promedio de 22 %.
- La variable "Tiempo de preparación Universitaria"- X_7 , resulta ser la segunda variable más importante, y tiene un efecto directamente proporcional, esto quiere decir, a más tiempo de preparación mayores serán los puntajes.
- La variable "Sexo del postulante"(X_3), resulta ser la tercera variable más importante y tiene un efecto inverso sobre el puntaje; esto es, existen mayores promedios en los postulantes de sexo masculino.
- La variable "Canal al que postulan"(X_1), resulta ser la cuarta variable más importante y tiene un efecto inverso sobre el puntaje del examen; y esto se interpreta como que los postulantes a los canales 2 y 3, tienen los puntajes más altos.
- La variable "Tipo de preparación Pre-Universitaria" (X_4), también resulta ser algo importante, y tiene un efecto positivo, indicando que el postulante que se prepara en las academias pre-universitarias, tendrán puntajes más altos.
- Los gráficos de la "señal ridge", revelan que existe una inestabilidad marcada en la estimación de los coeficientes de regresión, por lo que se recomienda usar la técnica de regresión ridge para encontrar una ecuación, si el propósito es la predicción.

REFERENCIAS BIBLIOGRÁFICAS

- HOCKING, R.R.**(1976) "The analysis and selection of variables in linear regression". Biometrics Vol. 32, Pags. 1-48.
- HOERL, A.E and KENNARD, R.W.** (1970) "Ridge regression: Biased estimation for non-orthogonal problems". Technometrics Vol. 12, Pags. 55-67.
- ORDOÑEZ, M.A.**(1987) "Regresión Cresta en la Selección de variables en modelos de Regresión Lineal Múltiple". Tesis para optar el título de Ingeniero Estadístico. en la Univ.Nac. Agraria La Molina. Lima-Perú.