

## Metodología para estadística inferencial y segmentación de la base de datos del Proyecto SANREM

### Análisis estadístico inferencial de las variables priorizadas en base a la relación de los módulos de la encuesta y a las zonas de trabajo.

#### Descripción de factores

Las variables consideradas como factor de clasificación de la zona de estudio son:

Factor	Nivel	Descripción
Zona "A"	4	1. Alta
		2. Intermedia
		3. Baja
		4. Externa al Proyecto
Comunidad "B"	18	1. Toralapa baja
		2. Toralapa alta
		3. Uchuchi cancha
		4. Waylla pujru
		5. Ch'aki khocha
		6. Caña cota
		7. 1ro de marzo
		8. Boqueron grande
		9. Chaupi rancho
		10. Suraj mayu
11. Damy rancho		
12. Cebada jich'ana		
13. Sankayani alto		
14. Boquerón alto		
15. Villa san Isidro		
16. Koari alto		
17. Kayarani		
20. Pairumani		

La información levantada tiene diferente unidades muestrales por comunidad y zona.

#### Modelo estadístico

Las variables cuantitativas (medibles) fueron procesadas con el modelo lineal generalizado bajo el supuesto de distribución gamma y poisson (Montgomery, 2003; Kachman, 2000) debido a que la mayoría de las variables no cumplen los supuestos  $(\sim \text{DNII}(0, \sigma_1^2))$  que exige las herramientas paramétricas (análisis de varianzas), por tanto el modelo base fue:

$$\eta_{ijk} = \eta + \tau_i + \alpha_{jk}$$

donde:

i= 1, 2, ... de a de zona

j= 1, 2, ... de b de comunidad

k= 1, 2, ... de r de unidad muestral "familia encuestada"

$\eta_{ijk}$ : Valor generalizado de una variable estudiada de la k-esima unidad muestral ubicada en la j-esima comunidad e i-esima zona

$\eta$ : Constante común

$\tau_i$ : Efecto fijo de la i-esima zona

$\alpha_{jk}$ : Efecto fijo de la j-esima comunidad dentro la i-esima zona

Las variables cualitativas (categóricas) fueron procesadas con el modelo lineal generalizado bajo el supuesto de distribución binomial y multinomial (Montgomery, 2003; Kachman, 2000)

El objetivo de esta herramienta estadística es determinar la existencia de diferencias entre zonas y entre comunidades dentro cada zona sobre las variables analizadas una por una. Este criterio responde al objetivo planteado en el Proyecto. Esta información también permitirá conocer la situación de las zonas y las comunidades.

Las variables con las herramientas estadísticas mencionadas serán procesadas con el programa estadístico SAS v8 (SAS Institute Inc. 2000)

**Brindar información sobre limitaciones y potencialidades identificadas en las comunidades de la Subcuenca del río Jatun Mayu para su utilización en el diseño de innovaciones tecnológicas.**

La nueva base de datos está conformada por variables priorizadas (en lo posible cuantitativas) de los 24 módulos y con 388 observaciones (familias).

El uso de la herramienta estadística multivariada Análisis de Componentes principales (Pla L, 1986; Hair, Anderson, Tatham y Black. 1999), pretende generar nuevas variables que puedan expresar la información de toda la base de datos original. Estas nuevas variables (componentes principales) permitirá identificar variables que tienen influencia positiva, negativa o intermedia (limitaciones y/o potencialidades), también se ubicará las familias con estas limitaciones o potencialidades. Además el análisis componentes principales coadyuvará eliminar algunas variables originales que aportan poca información.

La herramienta estadística Análisis Tipológico o Agrupación Jerárquica permitirá clasificar a las familias en grupos que sean lo más homogéneas en lo posible y los grupos sean distintos entre ellos (Pedert, Sagnier y Camp. 2003; Hair, Anderson, Tatham y Black. 1999). Éste análisis se ejecutará con la nueva base de datos (descrita anteriormente).

Con ambas herramientas se segmentará familias con debilidades y/o potencialidades expresadas con las variables de la base de datos, recordemos que estas variables comprenden sobre el manejo tecnológico de los cultivos, ingresos, beneficios y otras.

Las variables cualitativas también serán priorizadas en una segunda nueva base de datos, similar al criterio anterior se utilizará las herramientas estadísticas Análisis de correspondencia o factorial y Análisis tipológico (Pedert, Sagnier y Camp. 2003; Hair, Anderson, Tatham y Black. 1999;).

Las variables con las herramientas estadísticas mencionadas serán procesadas con el programa estadístico SAS v8 (SAS Institute Inc. 2000)