QUIPO 2.0

Flavio Lozano Isla

2017-04-16

# Bienvenidos

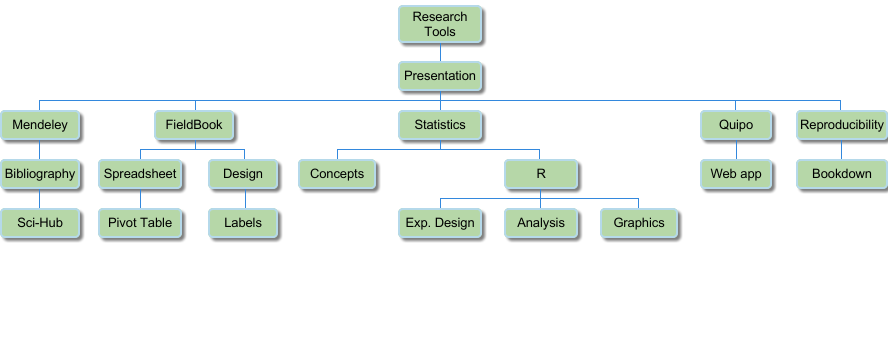


Quipu 2.0 es un guia para ayudar en el manejo y organización de diferentes tipos de experimentos en las areas de las ciencias agraria y biologicas con uso del software estadístico R.

# Por que leerlo?

Si eres un persona que tiene curiosidad de comenzar una investigacion y no sabes como inciar tu parte experimental, esta guia sera de gran ayuda. A traves de este curso se ira describiendo en forma continua y de manera sensilla la metodologia de la investigacion experimental que puede ser extendidas a otras áres de estudio, con herramientas de aceso libre que seran de gran ayuda. Muchas de estas herramientas tienen como fundamento la ejucucion de codigos, pero no se asusten, ya que se transmitira los conocimientos de una manera facil con un lenguaje sencillo.

# Organigrama de temas



# Syllabus

## Datos generales

**Programa :** Herramientas para la Investigación Científica

**Requisito :** Métodos estadísticos

**Duración :** 20 horas

**Requerimiento :**

* Laptop
* Gmail account (recomendado)

## Justificación

El desarrollo tecnológico e industrial de los países se basa en la investigación en las áreas de ciencias básicas y aplicadas. Por lo que es importante saber las herramientas para el planeamiento, manejo, análisis e interpretación de información colectada durante los procesos de investigación. Una de las mayores deficiencias para el desarrollo de la investigación es el desconocimiento de dichas herramientas que dificultan el proceso de investigación generando colecta errada de datos e inferencias no acordes a la realidad. Actualmente existen muchas herramientas para el planeamiento y análisis de datos, cuyos costos son muy elevados e inaccesibles para muchos estudiantes y profesionales. En este curso se harán uso de herramientas basados en código libre con amplia aplicación en diferentes ramas de las ciencias que faciliten los procesos de investigación.

## Objetivos

* Aprender la planificación y manejo de experimentos bajo distintos diseños estadísticos (DCA, DBCA, FACTORIAL).
* Aprender el uso de herramientas para el análisis estadístico con el software R.
* Comprender los análisis estadísticos e interpretación de resultados estadísticos.
* Desarrollo de gráficos estadísticos en R.
* Usar herramientas para facilitar la redacción científica.

## Cronograma y contenido de curso

|  |  |
| --- | --- |
| Horas | Contenido |
| 4 | Presentación y requisitos del curso |
|  | Lenguaje R y uso de librerías |
|  | Tipos de variables |
| 4 | Importación de base de datos a R |
|  | Manipulación, organización y visualización de datos |
|  | Resumen de base de datos |
| 4 | Metodología de investigación científica |
|  | Planteamiento y diseño de libro de campo |
|  | Colecta e importación de datos |
| 4 | Conceptos basicos de estadistica |
|  | Modelos y diseños experimentales |
| 4 | Gráfica de resultados de análisis estadísticos |
|  | Análisis multivariado |
|  | Presentacion de la aplicación fieldbook |

## Softwares

* [R](https://cran.rstudio.com/)
* [R Studio](https://www.rstudio.com/products/rstudio/download/)
* [Mendeley](https://www.mendeley.com/)
* [Github desktop](https://desktop.github.com/)

## Requerimientos técnicos

* Conexión a internet
* Enchufes para cargar laptops
* Proyector Multimedia
* Pizarra

# Introducción

## Método científico

Siempre nos estamos haciendo preguntas de las cosas que suceden a nuestro alrededor, para responder a esas preguntas de forma sistematica, se hace uso del metodo cientifico. El metodo cientifico se incia con una observacion de algo que nos llamas la atencion, esto genera preguntas. Con las preguntas que nos realizamos podemos sugerir hipotesis de por que esta ocurriendo dicho fenómeno, por lo que se recurre a un forma de comprobar la hipotesis a traves de un experimento, donde rechazaremos o damos como cierta nuestras hipotesis anteriormente sugeridas, dando como resultado informacion que nos lleva a conclusiones del trabajo, muchas veces esta informacion concluye en esta fase, ya que pocos documentan los resultados. Si los resultados son publicados se generan nuevas ideas o descubrimientos, lo cual puede generar nuevas preguntas para otras personas que deseen seguir explorando dicho tema y es ahi donde se genera el conocimiento, Figura 1.



Figure 1 Pasos del método científico

## La investigacion y la publicacion

La investigacion cientifica y la publicación del artículo científico son dos actividades intimamente relacionadas. Algunas personas creen que la investigación termina cuando se obtienen los resultados, cuando estos se analizan, cuando se entrega el informe del trabajo o cuando la investigación se presenta en una reunión profesional. Sin embargo, la investigación científica formal y seria termina con la publicación del artículo científico; solo así tu contribución pasará a formar parte del conocimiento científico. Algunas personas van más lejos y sugieren que la investigación termina cuando el lector entiende el artículo; es decir, que no basta con publicar el trabajo, también es necesario que la audiencia entienda claramente su contenido (Mari Mutt [2005](#ref-mari2005manual)). Un analisis brillante resulta inutil si los resultados no son comunicados en lenguaje sencillo y claro (Gonick y Smith [1993](#ref-Gonick1993)). Para escribir un buen artículo científico tienes que conocer y practicar los tres principios básicos de la redacción científica: precisión, claridad y brevedad.

## Problemas comunes en la investigación

Uno de los principales porblemas que pude observar a disitintos niveles, tanto en estudiantes como en investigadores, es la falta de organizacion y orden no sólo en las bases de datos, si no en la forma de ejecucion de sus experimentos. Un experimento al ser parte de un plan de trabajo debe ser planificado con anticipacion y tratar de reducir al minimo las posibles los errores o sincronizacion de los tiempos en las evaluaciones y toma de dato. Sumado a este problema los investigadores no suelen tener un diario a para hacer las anotaciones de sus trabajos, esto junto a la mala colecta de la informacion, que usualmnete son en hojas de calculo excel, genera un situacion cada vez más engorroza en la etapas posterioes como el analisis de datos y presentacion de los resultados.

# Softwares y aplicaciones

Los software que se haran uso en este guia, son de distribucion gratuitas o de codigo abierto, con esto se quiere fomentar la reducción del uso de software pirata y los costos de licencias de softwara de pago.

## Software estadistico R

En el año 1992, los investigadores Ross Ihaka y Robert Gentleman, ambos de la Universidad de Auckland, Nueva Zelandia, desarrollaron un software para el analisis estadistico y le atribuyeron el nombre de software R, el cual fue basado en el lenguaje de programacion S, inicialmente desarrollado en los laboratorios de *AT&T Bell* a mediados del año 1970 (Calenge [2006](#ref-calenge2006package)).

Cuando R. Ihaka e R. Gentleman lanzaron el software R, le atribuyeron funcionalidad de ser de código abierto (Ihaka y Gentleman [1996](#ref-Ihaka2012)), lanzando la plataforma "Comprehensive R Archive Network (CRAN)", que es el nucleo donde se hospeda la mayoria de los paquetes del software R. El CRAN es una comunidad cada vez creciente y es el principal repositorio de material relacionado a R (Hornik y Leisch [2002](#ref-Hornik2002)).

Desde su creacion, R paso a ser un lenguaje de programacion y ambiente para el analisis estadistica computacional y generacion de graficos. El software R da una amplia variedad de analisis estadisticos, tales como, modelaje lineal y no lineal, testes estadisticos clasicos, analisis de series temporales, clasificacion, cluster y técnicas graficas, todas de alta reporductivilidad. O R fue desarrollado en un entorno de lenguaje de programacion legitima, y permite a los usuarios adicionar nuevas funcionalidades de acuerdo a sus necesidades. Las funcionalidades de R pueden ser extendidas a traves de paquetes para desarrollar ciertas funciones determinadas. Actualamente existen 8 paquetes basicos fornecidos por la distribucion de R y otras tantas estan disponibles a traves del CRAN, con servidores en la internet que cubren una amplia gama de herramientas estadisticias modernas e R e outroas tantaos (R Core Team [2017](#ref-R-base)).



Figure 2 Por que usar R?

## R studio

RStudio es un entorno de desarrollo integrado (IDE) para R. Incluye una consola, editor de resaltado de sintaxis que soporta la ejecución directa de código, así como herramientas para graficar, historial, la depuración y la gestión del espacio de trabajo. RStudio está disponible en codigo libre y ediciones comerciales y se ejecuta en el escritorio (Windows, Mac y Linux).

"RStudio makes R easier to use. It includes a code editor, debugging & visualization tools. RStudio (2017)"

**Principales caracterisitcas**

* Integracion a la ayuda y documentación de R.
* Administrar fácilmente múltiples directorios de trabajo mediante proyectos.
* Navegador de espacio de trabajo y visualizador de datos.
* Ejecutar código R directamente desde el editor fuente.
* Saltar rápidamente a las definiciones de funciones.
* Extensas herramientas de desarrollo de paquetes.
* Redacción de documentos cientificos.



Figure 3 Interface de RStudio

## Mendeley

Mendeley es un gestor de referencia gratuito y una red social académica que ayuda en la organizar de las investigaciones, colaborar con otros en línea y descubrir las últimas investigaciones.

**Principales caracterisitcas**

* Generar automáticamente bibliografías.
* Colaborar fácilmente con otros investigadores en línea.
* Importar fácilmente documentos de otro software de investigación.
* Encuentrar documentos relevantes basados en lo que está leyendo.
* Acceder a sus documentos desde cualquier lugar en línea.
* Leer documentos sobre la marcha, con las aplicaciones para dispositivos moviles.

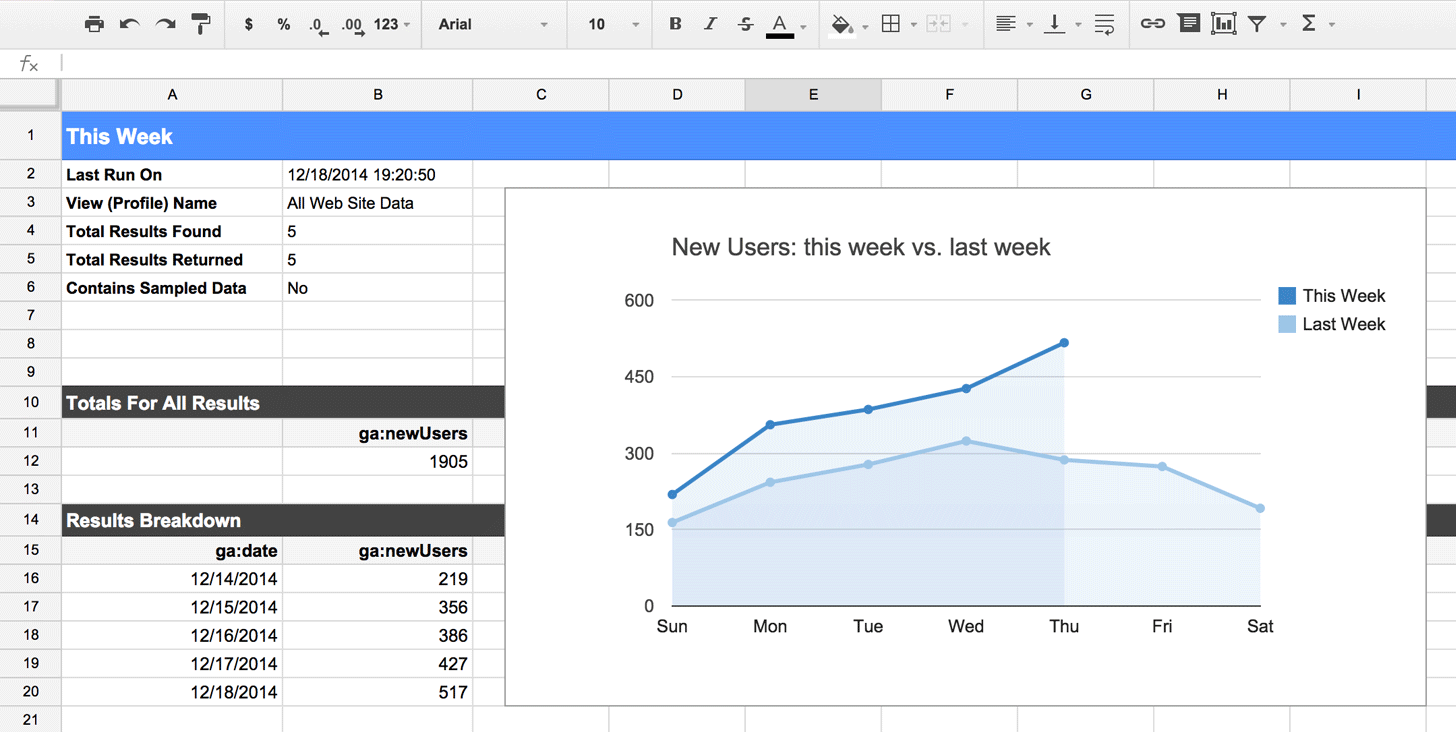


## Google spreadsheets

Google spreadsheet es un servicio vía web de hojas de cálculo, realizado en tecnología *AJAX*. Con ella se puede realizar la mayoría de las funciones que dejan las aplicaciones de hojas de cálculos de los programas ofimáticos (microsoft office), como realizar operaciones entre celdas con diferentes tipos de funciones (matemáticas, financieras, lógicas, de fechas, de búsquedas, estadísticas, con cadenas e informativas), ordenar columnas, manejar diferentes hojas dentro de cada fichero, manejas ficheros del tipo *xlsx* y *csv*, etc.

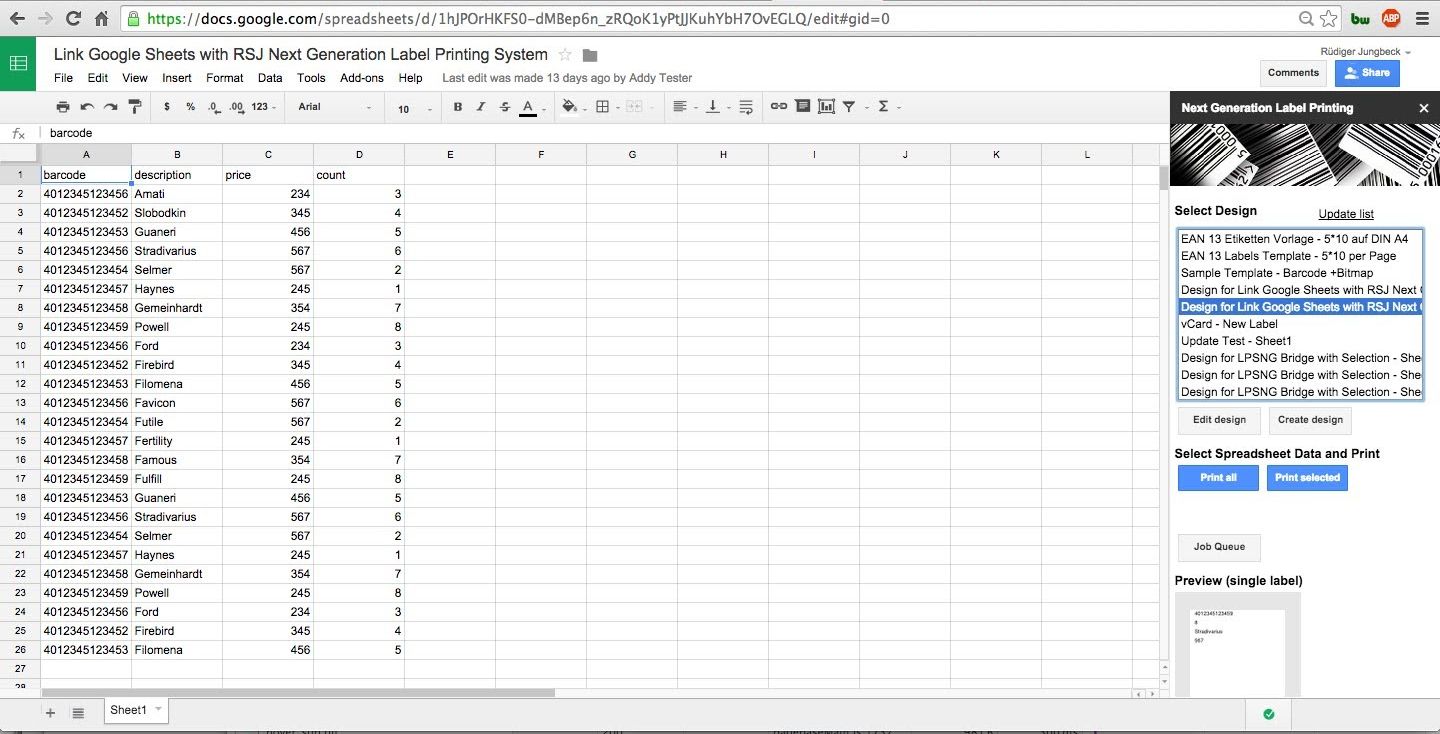
**Principales usos**

1. Organizacion y colecta de datos.
2. Tablas dinamicas.
3. Generacion de etiquetas.



## Next Generation Label Printing System

**RSJ Next Generation Label Printing System** es una aplicacion para facilitar la creación e impresión de etiquetas que contienen datos suministrados directamente en su navegador y como componente de las hojas de calculo de google spreadsheet.



# Planeamiento de experimentos

El planeamineto de un experimento es la parte crusial para definir el éxito de un trabajo experimental, mismo asi cuando se realice un plan de trabajo, acontencen dudas e imprevistos que van aconteciendo. Para ayudar en esta tarea introduciremos un termino **fieldbook** que significa "libro de campo" y sera realizado en una hoja de calculo, el cual contiene los datos minimos que son necesarios para realizar tu experimento, de esta manera ahorrar tiempo en el procesamiento de los datos, ya que muchas veces no es el analisis el que consume el tiempo del investigador, si no la reorganizacion de los datos colectados durante el experimento.

Los coputadores presentan una forma especial de leer los datos, ya que muchas veces organizamos estos de la manera que no es mas facil entender a nosotros, lo que genera que al final debemos reorganizar todo, para que el computador pueda entenderlo.

## Fieldbook

El libro de campo esta constituido por cuatro partes esenciales (las cuales pueden ser aumentadas de acuerdo a las necesidades (ie. Diseño de campo, informacion metereologica, planilla de riego, etc) para cualquier tipo de experimento.

1. Informacion. Se introducira la informacion del proyecto, como lugar de ejeccuon, tipo de diseño experimental, colaboradores, georeferencia, fecha de inicio y tambien los objetivos del trabajo.
2. Variables. Se realizará una lista de las variables a evaluar durante el experimento, tambien se adicionara las abreviaturas de las variables las cuales iran como encabezados en la hoja de diseño estadistico.
3. Diseño estadistico. Se introducira el tipo de diseño experimental (ie. DCA, DBCA, DCL) los cuales pueden ser generados con el paquete *agricolae* (de Mendiburu [2016](#ref-R-agricolae)).
4. Calendario. Se generará el calendario de evaluaciones y actividades para poder organizar las acciones durante el experimento (ie. uso de materiales, pretamos de equipos, disponibilidad de espacios, etc)
5. Diario. Se recolectara informacion de lo que va aconteciendo en el experimento, se registrara el dia y las observaciones, asi como comentarios personales.

Typically a scientist or anyone who works with data will end up with several files with different copies of the data and several files with intermediate steps. If you are working with data, there should have been a time where you had a file with name data, but now you probably have:

StatTools, 2014 datasharing , 2014

* data
* data\_V01
* data\_V02
* data\_V02\_1
* data\_DDMMYY
* lastdata
* finaldata
* veryfinaldata
* and some others.

## Colecta de datos

### Etiquetas

El etiquetado de los datos es una actividad que siempre se deja de lado, y muchas veces por el trabajo que genera realizarla en la unidades experimentales. Pues piensalo 2 veces antes de dejar esta actividad para el final, un etiquetado adecuado, facilitara la evaluacion y toma de datos en todas las instancias del proyecto.

### Evaluacion de las variables

Independientemente del tipo de diseño que hayas elegido, este será utilizado para **analizar** los datos, en cuanto a la colecta de datos, la informacion sera tomada e inserida al libro de campo para cada unidad experimental.

# Clases

# Cursos y participantes

## Univeridad Nacional Agraria la Molina (UNALM)

**Organizador:** Facultad de Agronomía.

**Lugar:** UNALM, Lima, Perú.

**Fecha:** 27, 28, 30 Junio y 1, 2 de Julio 2016.

**Horas:** 20 horas.

**Profesores:**

* Omar Benites Alfaro
* Flavio Lozano Isla

Curso dirigido a los profesores y alumnos de la facultad de Agronomia de la UNALM con soporte financiero de Consorcio de Universidades Flamencas de Bélgica (VLIR) con el objetivo de fortalecer y mejorar la calidad de la educación en la UNALM.

Table 1 Participantes del curso. Junio 2016, UNALM

|  |  |
| --- | --- |
| PARTICIPANTES | EMAIL |
| ALEJANDRO RISCO MENDOZA | [ariscom22@gmail.com](mailto:ariscom22@gmail.com) |
| ALFREDO ALBERTO BEYER ARTEAGA | [abeyer@lamolina.edu.pe](mailto:abeyer@lamolina.edu.pe) |
| ANDRES VIRGILIO CASAS DIAZ | [cda@lamolina.edu.pe](mailto:cda@lamolina.edu.pe) |
| ANGEL ALFONSO PALOMO HERRERA | [fonz@lamolina.edu.pe](mailto:fonz@lamolina.edu.pe) |
| BRAULIO LA TORRE MARTINEZ | [braulio@lamolina.edu.pe](mailto:braulio@lamolina.edu.pe) |
| DAVID SARAVIA NAVARRO | [davidsaravian@gmail.com](mailto:davidsaravian@gmail.com) |
| EDGARDO VILCARA CARDENAS | [eavilcara@lamolina.edu.pe](mailto:eavilcara@lamolina.edu.pe) |
| EDU TORRES SANCHEZ | [20101082@lamolina.edu.pe](mailto:20101082@lamolina.edu.pe) |
| GONZALO CABRERA ZARATE | [gonzalocabrera.91@gmail.com](mailto:gonzalocabrera.91@gmail.com) |
| GUILLERMO AGUIRRE YATO | [gaguirre@lamolina.edu.pe](mailto:gaguirre@lamolina.edu.pe) |
| HUGO HUANUQUEÑO COCA | [ehh.coca@lamolina.edu.pe](mailto:ehh.coca@lamolina.edu.pe) |
| JORGE ALBERTO COHEN MORALES | [jorge\_cohen\_m@hotmail.com](mailto:jorge_cohen_m@hotmail.com) |
| JORGE LUIS TEJADA SORALUZ | [jorgetejada@lamolina.edu.pe](mailto:jorgetejada@lamolina.edu.pe) |
| JUAN CARLOS JAULIS | [jjaulis@lamolina.edu.pe](mailto:jjaulis@lamolina.edu.pe) |
| JUAN FLORES TORRES | [gerardo@lamolina.edu.pe](mailto:gerardo@lamolina.edu.pe) |
| KAREN MINCHOLA VILLACORTA | [20110049@lamolina.edu.pe](mailto:20110049@lamolina.edu.pe) |
| LILIANA MARIA ARAGON CABALLERO | [lili@lamolina.edu.pe](mailto:lili@lamolina.edu.pe) |
| MARLENE AGUILAR | [maguilarhe@lamolina.edu.pe](mailto:maguilarhe@lamolina.edu.pe) |
| MEDALI HEIDI HUARHUA | [medalihuarhua@lamolina.edu.pe](mailto:medalihuarhua@lamolina.edu.pe) |
| PATRICIA QUIROZ | [patriciaquiroz@lamolina.edu.pe](mailto:patriciaquiroz@lamolina.edu.pe) |
| PEDRO PABLO GUTIERREZ VILCHEZ | [pablogv0921@gmail.com](mailto:pablogv0921@gmail.com) |
| RUBY VEGA RAVELLO | [rvega@lamolina.edu.pe](mailto:rvega@lamolina.edu.pe) |
| SARITA MORENO | [saritamoreno@lamolina.edu.pe](mailto:saritamoreno@lamolina.edu.pe) |
| SOFIA JESUS FLORES VIVAR | [sofia.flores.vivar@gmail.com](mailto:sofia.flores.vivar@gmail.com) |

Agradecimientos a las Profesora Ing. Liliana Aragon por el apoyo y confianza para la organizacion del curso, al Ing. Pablo Gutierrez por organizar y ser el promotor del evento.

## Univeridad Nacional Agraria la Molina (UNALM)

**Organizador:** Circulo de Investigacion en Fitomejoramiento (CIF).

**Lugar:** UNALM, Lima, Perú.

**Fecha:** Setiembre de 2014.

**Horas:** 25 horas.

**Profesores:**

* Flavio Lozano Isla

Curso libre impartido a los alumnos voluntarios de las distintas facultades de la UNALM.

Table 2 Participantes del curso. Setiembre 2014, UNALM

|  |  |
| --- | --- |
| PARTICIPANTES | EMAIL |
| CARLOS JULIAN MESTANZA NOVOA | [julian.cmn@gmail.com](mailto:julian.cmn@gmail.com) |
| DIEGO ACUÑA | [leo\_aioroz@hotmail.com](mailto:leo_aioroz@hotmail.com) |
| ELENA NUÑEZ | [elenuli.062@gmail.com](mailto:elenuli.062@gmail.com) |
| JOSE FRANCO VILLAFUENTE | [20080925@lamolina.edu.pe](mailto:20080925@lamolina.edu.pe) |
| MARCELO GAMBOA ESTRADA | [mgamboaestrada@gmail.com](mailto:mgamboaestrada@gmail.com) |
| MIGUEL MATICORENA QUISPE | [maticorenaquispe@gmail.com](mailto:maticorenaquispe@gmail.com) |
| PAMELA CASTRO | [20091043@lamolina.edu.pe](mailto:20091043@lamolina.edu.pe) |
| RODRIGO PEZUA | [20101062@lamolina.edu.pe](mailto:20101062@lamolina.edu.pe) |
| ROMINA CAMACHO AÑAZCO | [romina.camacho@gmail.com](mailto:romina.camacho@gmail.com) |
| WILLIAMS SERNAQUE | [williams\_9114@hotmail.com](mailto:williams_9114@hotmail.com) |
| YONY CALLOHUARI | [ycallohuari@gmail.com](mailto:ycallohuari@gmail.com) |

Agradecimientos al Profesor Ing. Andres Virgilio Casa por ayudar en la organizacion y facilitar el uso de los salones de la universidad para poder impartir las clases.

## Univeridad Nacional Agraria la Molina (UNALM)

**Organizador:** Circulo de Investigacion en Fitomejoramiento (CIF).

**Lugar:** UNALM, Lima, Perú.

**Fecha:** Julio de 2014.

**Horas:** 25 horas.

**Profesores:**

* Flavio Lozano Isla

Curso dirigido a los alumos de los distintos circulos de investigacion de la UNALM y otras facultades para fomentar y ayudar a los participantes en la investigacion y desarrollo de sus trabajos de tesis.

Table 3 Participantes del curso. Julio de 2014, UNALM

|  |  |
| --- | --- |
| PARTICIPANTES | EMAIL |
| CARMEN DEL PILAR LIVIA TACZA | [20080931@lamolina.edu.pe](mailto:20080931@lamolina.edu.pe) |
| ERIK NIKOL MUÑOZ CABALLERO | [erikmunozcaballero@gmail.com](mailto:erikmunozcaballero@gmail.com) |
| FIDEL HAUSSEN ROJAS IRURI | [20110072@lamolina.edu.pe](mailto:20110072@lamolina.edu.pe) |
| HARUMI HAMAMOTO KOHAMA | [hamamoto.harumi@gmail.com](mailto:hamamoto.harumi@gmail.com) |
| IVAN VALDIVIEZO ALIAGA | [20110993@lamolina.edu.pe](mailto:20110993@lamolina.edu.pe) |
| JEAN P. SAUX CASANAVE | [20090067@lamolina.edu.pe](mailto:20090067@lamolina.edu.pe) |
| JIMMY RENE GOMEZ CARRION | [20070033@lamolina.edu.pe](mailto:20070033@lamolina.edu.pe) |
| JOAQUIN A. SALINAS ANGELES | [20120084@lamolina.edu.pe](mailto:20120084@lamolina.edu.pe) |
| JONATHAN VEGA | [jonathan\_2610@hotmail.com](mailto:jonathan_2610@hotmail.com) |
| JOSE J. TOLEDO CHOQUEHUANCA | [20110079@lamolina.edu.pe](mailto:20110079@lamolina.edu.pe) |
| LILIANA HIROMI NAKANDAKARI | [nhiromi4@hotmail.com](mailto:nhiromi4@hotmail.com) |
| RONALD VALERIANO | [valerianuz@hotmail.com](mailto:valerianuz@hotmail.com) |
| ROSAURA LAURA VILA | [rosauralaura@gmail.com](mailto:rosauralaura@gmail.com) |
| ROSSANA PORRAS JORGE | [rossanaporrasjorge@gmail.com](mailto:rossanaporrasjorge@gmail.com) |
| SEBASTIAN IAGO DAVIS BAYLY | [sdavisbayly@gmail.com](mailto:sdavisbayly@gmail.com) |
| SILVIA SIFUENTES AMEZ | [20090484@lamolina.edu.pe](mailto:20090484@lamolina.edu.pe) |

Agradecimientos al Dr. Raul Blas por su ayuda y facilidad para la organizacion del curso. A Rosaura Laura y Sebastian Davis por ser los promotores y organizadores a traves del CIF y el Laboratorio de Biotecnologia de la UNALM.

# Referencias

Calenge, C. 2006. The package adehabitat for the R software: a tool for the analysis of space and habitat use by animals. Ecological modelling 197(3): 516-519.

de Mendiburu, F. 2016. Agricolae: Statistical Procedures for Agricultural Research

Gonick, L; Smith, W. 1993. The Cartoon guide to Statistics. Ed. HarperCollins. First New York, HarperCollins, p.240.

Hornik, K; Leisch, F. 2002. Vienna and R: Love, marriage and the future. s.l., Citeseer.

Ihaka, R; Gentleman, R. 1996. R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics 5(3): 299-314.

Mari Mutt, JA. 2005. Manual de redacción científica. Universidad de Puerto Rico 2005.

R Core Team. 2017. R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing.