INKAVERSE

Tarpuy: Tidy Data y Planes Experimentales Interactivos

Flavio Lozano-Isla

Problemas

- Proyectos sin información mínima
- Incorrecta organización de la información colectada
- Trabajos almacenados sin analizar
- Dificultad para el uso/licencia de softwares estadísticos

¿Cómo hacer pastas?

Slide 4

¿Por qué es importante la investigación?

➡ Fideos en base de legumbres

	F	L	BG	C
Color	L* = 38.7 a* = 9.1 b* = 19.5	L* = 27.2 a* = 1.2 b* = 4.2	L* = 50.5 a* = 5.4 b* = 28.2	L* = 55.1 a* = 3.5 b* = 37.6

F: faba, L: lentil, BG: black-gram and C: commercial pasta.

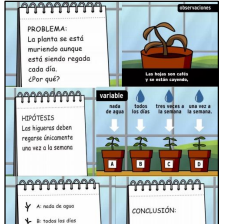Laleg, Karima, et al. "Structural, Culinary, Nutritional and Anti-Nutritional Properties of High Protein, Gluten Free, 100% Legume Pasta." *PLOS ONE*, vol. 11, no. 9, Public Library of Science, Sept. 2016, p. e0160721. *PLoS Journals*, doi:10.1371/journal.pone.0160721.

Quinoa pasta (cooked); Q–A (quinoa flour), QQE10-A (quinoa flour, extruded quinoa flour 10%) QQE20-A(quinoa flour, extruded quinoa flour 20%) QQE35-A (quinoa flour, extruded quinoa flour 35%)

Linares-García, Laura, et al. "Development of Gluten-Free and Egg-Free Pasta Based on Quinoa (*Chenopodium quinoa* Willd) with Addition of Lupine Flour, Vegetable Proteins and the Oxidizing Enzyme POx." *European Food Research and Technology*, vol. 245, no. 10, Oct. 2019, pp. 2147–56. *Springer Link*, doi:10.1007/s00217-019-03320-1.

Slide 5

¿Cómo no hacer nada?

Slide 6

¿Por qué es importante el método científico?

- Rigor científico
- Replicabilidad
- Objetividad
- Comparativos

¿Dónde puedo recolectar mis datos?

Excel vs Google Sheets **Ventajas**

- Trabajo colaborativo
- Historial de versiones
- Es gratis
- Multidispositivos

¿Cómo organizar mis datos?

Variables

Observations

Recomendaciones

- NO unir celdas
- NO usar filtros
- Evitar el uso de caracteres extraños
- ⇒ i.e @, #, $
- Formato fechas
- ⇒ YYYY-MM-DD
- Usar abreviaciones en las cabeceras:
- ⇒ altura de planta → altp
- Aprender tablas dinámicas

Organización de Datos en Hojas de Cálculo

Data Organization in Spreadsheets

Karl W. Broman[a] and Kara H. Woo[b]

[a]Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison, WI; [b]Information School, University of Washington, Seattle, WA

ABSTRACT
Spreadsheets are widely used software tools for data entry, storage, analysis, and visualization. Focusing on the data entry and storage aspects, this article offers practical recommendations for organizing spreadsheet data to reduce errors and ease later analyses. The basic principles are: be consistent, write dates like YYYY-MM-DD, do not leave any cells empty, put just one thing in a cell, organize the data as a single rectangle (with subjects as rows and variables as columns, and with a single header row), create a data dictionary, do not include calculations in the raw data files, do not use font color or highlighting as data, choose good names for things, make backups, use data validation to avoid data entry errors, and save the data in plain text files.

Libro de campo: estructura

Pestañas   **Nombre de libro de campo**

1. Información
2. Lista de variables
3. Diseño experimental
4. Libro de campo
5. Croquis de experimento
6. Diario
7. Calendario
8. Presupuesto

- Localidad
- Año - Mes - Día
- Descripción

> Ejemplo

LA MOLINA 2020-12-22 (FERTILIZACIÓN MAÍZ)

Base de datos

Información   **Base de datos original**

Pestaña ⇒ info

- Título
- Autor
- Año
- Institución
- Documento
- Fuente
- Proveedor
- Tipo
- Resumen

Pestañas

info → información de los datos

original → datos originales

fb → datos organizados

Tablas dinámicas

- Correcta organización de la información
- Distintas formas de resumir la información
- Fácil generación de gráficos y tablas
- Primer paso para:
  - Eficiencia en manejo de la información
  - Programación

Libro de campo: problemas

un experimento → un archivo   **ejemplo**

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

Plan experimental (PLEX)

Esquematizar mi idea

1. Idea
2. Meta
3. Hipótesis
4. Fundamento
5. Objetivos
6. Plan