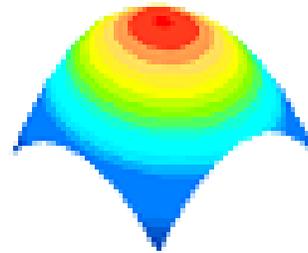
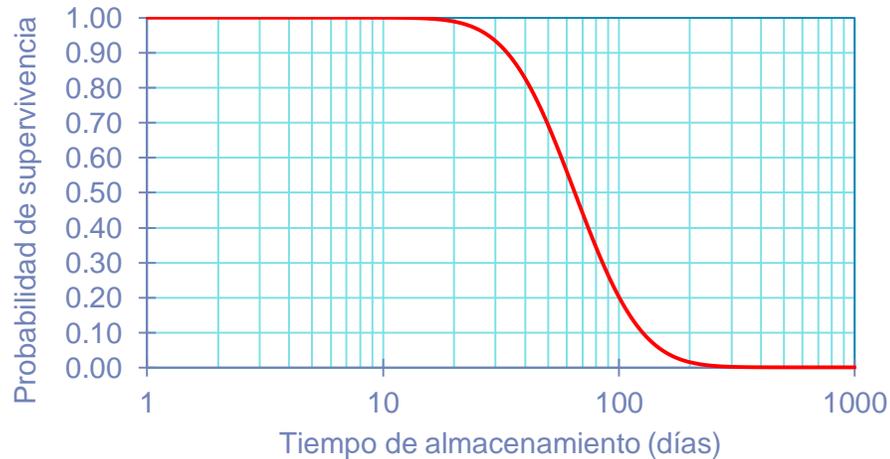
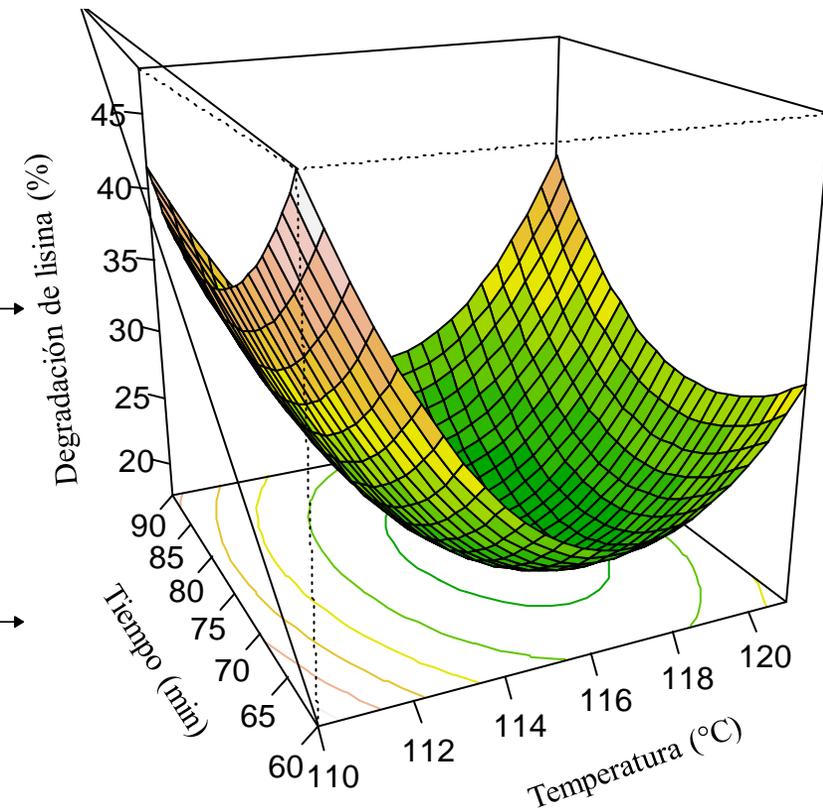
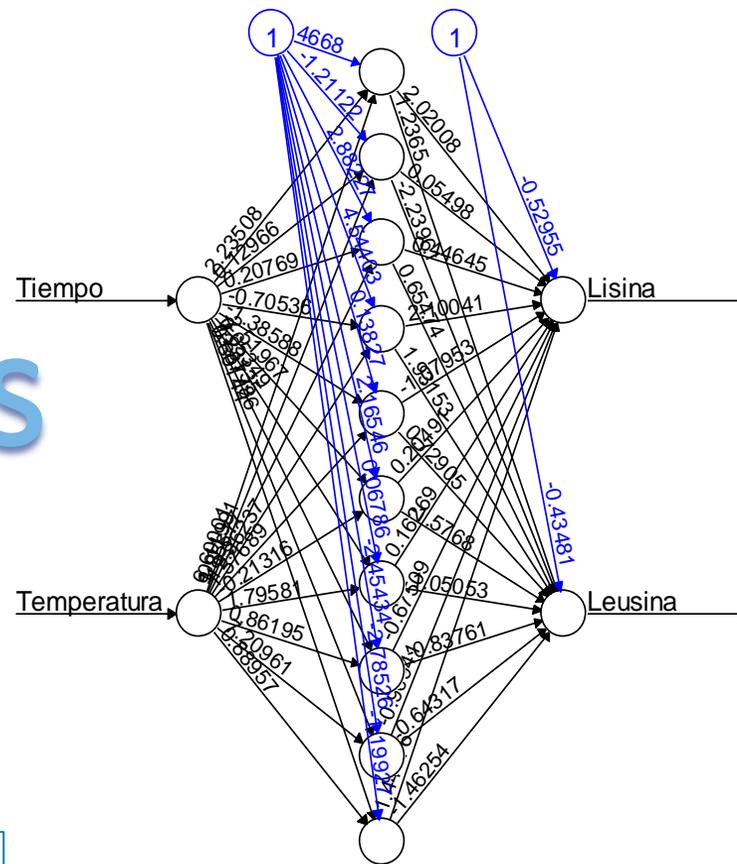


Taller online gratuito:

# Diseño de experimentos con



## Data Engineering

*Selección e Ingeniería*



# Expositor:



**Ingeniero en Industrias Alimentarias  
egresado de la Universidad Privada  
Antenor Orrego.**



UNIVERSIDAD NACIONAL  
DE TRUJILLO

**Maestro en Ciencias con mención en  
Estadística Aplicada.**



**Gerente General**

<http://dataengineeringperu.com/>



# Expositor:



## Nuestra historia

Todo comienzo es difícil dicen algunos conocedores; pues no fue difícil, pero si hubo mucho trabajo, esfuerzo y dedicación para hacer de esta idea una realidad y sobre todo romper paradigmas profesionales. Todo inicia el año 2011 con el vacío existente entre ingenieros y estadísticos (yo no te entiendo a ti y tú no me entiendes a mí... algo así) un gran dilema en la evaluación sensorial de alimentos y posterior procesamiento de datos.

El camino fue largo, aprendiendo y dando soluciones a problemas en procesamiento de datos haciendo uso de herramientas correctas de la estadística y aplicando la ingeniería para estar inmersos dentro del fenómeno en estudio; teniendo en claro desde nuestros inicios la "Mejora continua" y aplicando "Feedback" en nuestra organización; a finales de octubre del 2018 se consolida formalmente Data Engineering EIRL, para no sólo ser un servicio, sino una experiencia en el mundo del procesamiento de datos.



<http://dataengineeringperu.com/>

Ing. Mg. Jesús Alfredo Obregón Domínguez

Email: [gerencia@dataengineeringperu.com](mailto:gerencia@dataengineeringperu.com)

# Expositor:

CITE agroindustrial  
Chavimochic



Grupo  
**VIDA SOL**  
S.A.C.



**Innóvate** Perú



<http://dataengineeringperu.com/>

Ing. Mg. Jesús Alfredo Obregón Domínguez

Email: [gerencia@dataengineeringperu.com](mailto:gerencia@dataengineeringperu.com)

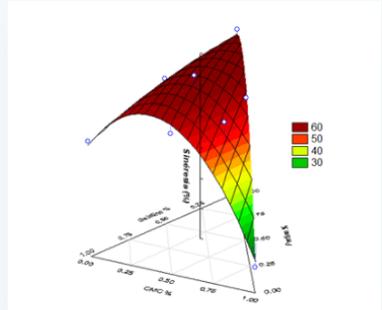
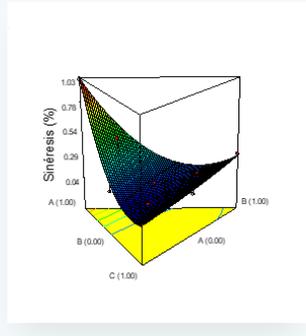
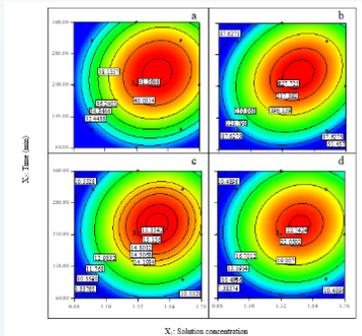
# Expositor:

## Nuestro aporte en estas investigaciones:

Osmotic pretreatment to assure retention of phenolics and anthocyanins in berry jams

Sinéresis, características reológicas y consistencia sensorial de salsa de alcachofa (*Cynara scolymus* L.)

Optimización mediante diseño de mezclas de sinéresis y textura sensorial de yogurt natural batido utilizando tres tipos de hidrocoloideos



Malting process as an alternative to obtain high nutritional quality quinoa flour

Native Andean potatoes (*Solanum tuberosum* L.): Phytonutrients in Peel, Pulp and Potato Cooking Water

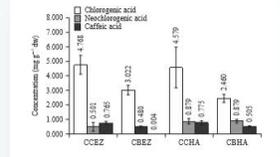
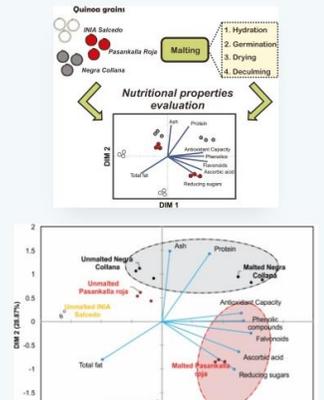
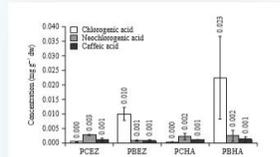


Fig. 1: Chlorogenic, neo-chlorogenic and caffeic acid in raw peel and cooked peel of Huagalina potato

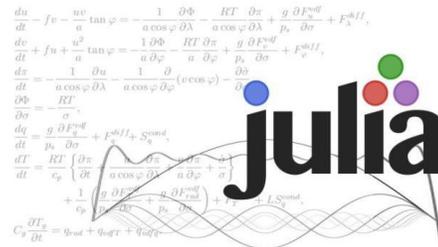
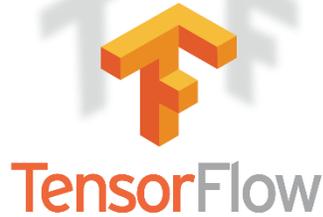


<http://dataengineeringperu.com/>

# Expositor:

## Áreas de interés

- Inteligencia artificial
- Machine learning
- Deep learning
- Estadística deportiva
- Diseño de experimentos
- Bioestadística
- Mercadotecnia
- Mejora continua
- Pruebas sensoriales de alimentos
- Determinación de vida útil
- Enfoque Bayesiano
- Six sigma



# ¿Qué es R?

---



# ¿Qué es R?

---

R es un lenguaje de programación utilizado para:

1. Manipulación de datos.
2. Cálculo.
3. Gráficos.



# Historia de R

---

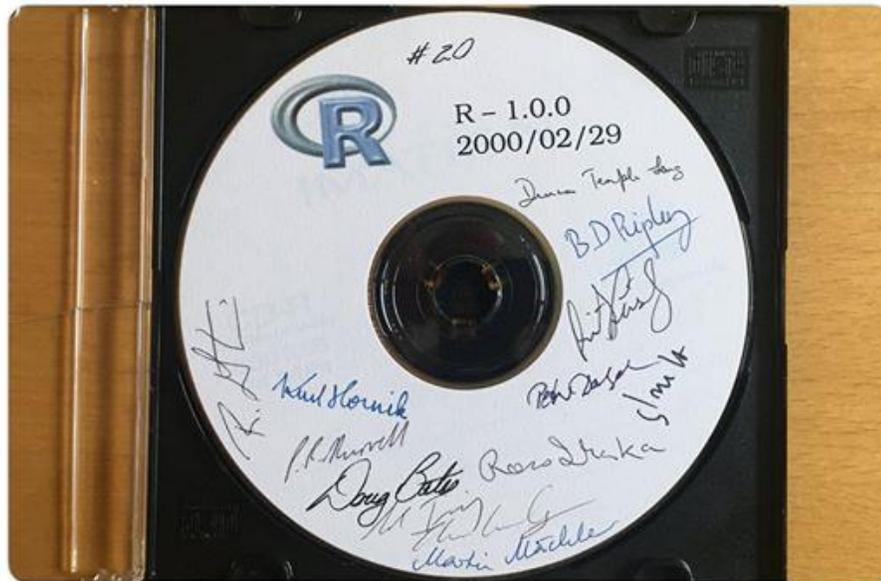


# Historia de R



Today is the 20th anniversary of the release of R 1.0.0.

Traducir Tweet



3:34 a. m. · 29 feb. 2020 · Twitter for iPhone

En 1991 en el Departamento de Estadística de la Universidad de Auckland en Nueva Zelanda, **Ross Ihaka** y **Robert Gentleman** crean **R** como un subdialecto de S e implementado su propio dialecto. El nombre de R es debido a la primera letra de sus creadores Ross y Robert. Anunciándolo en 1993 al público y en 1995 Martin Machler convence a Ross y Robert de usar la Licencia Publica General GNU haciendo a R como software libre.

# Por que aprender R?

---

1. R es gratuito. No necesitan piratear o comprar licencias de software comercial.
2. R tiene una comunidad académica detrás que provee una muy buena documentación en línea.
3. R crea gráficos de calidad superior.
4. Es compatible con 'todos' los formatos de datos (csv, xls, sav, sas. . . )
5. R es multiplataforma, es decir funciona en Mac, Windows o Linux).

# Por que aprender R?



[Home]

## Download

[CRAN](#)

## R Project

[About R](#)

[Logo](#)

[Contributors](#)

[What's New?](#)

[Reporting Bugs](#)

[Conferences](#)

[Search](#)

[Get Involved: Mailing Lists](#)

[Developer Pages](#)

[R Blog](#)

## R Foundation

[Foundation](#)

[Board](#)

[Members](#)

[Donors](#)

[Donate](#)

## The R Project for Statistical Computing

### Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

### News

- **R version 4.0.2 (Taking Off Again)** has been released on 2020-06-22.
- [useR! 2020 in Saint Louis has been cancelled](#). The European hub planned in Munich will not be an in-person conference. Both organizing committees are working on the best course of action.
- **R version 3.6.3 (Holding the Windsock)** has been released on 2020-02-29.
- You can support the R Foundation with a renewable subscription as a [supporting member](#)

### News via Twitter

 The R Foundation Retweeted

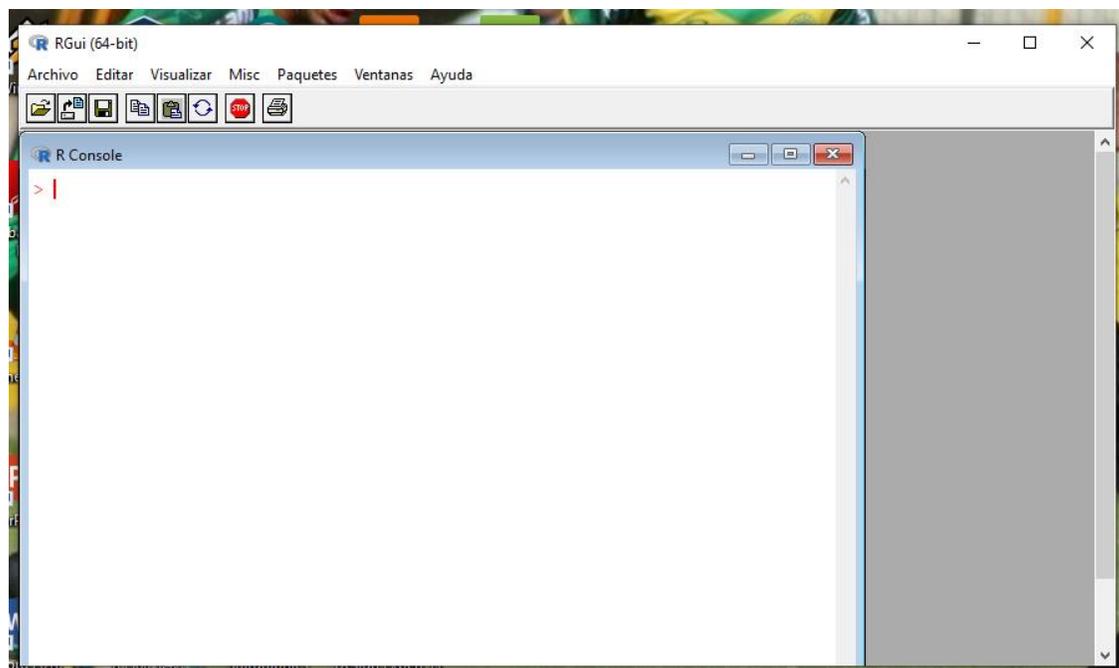
 **useR2020muc**  
@useR2020muc

To make it easier for you to know about times and dates of the #useR2020 live events, we've integrated a calendar on our website which should show events in your time zone. #useR2020 r

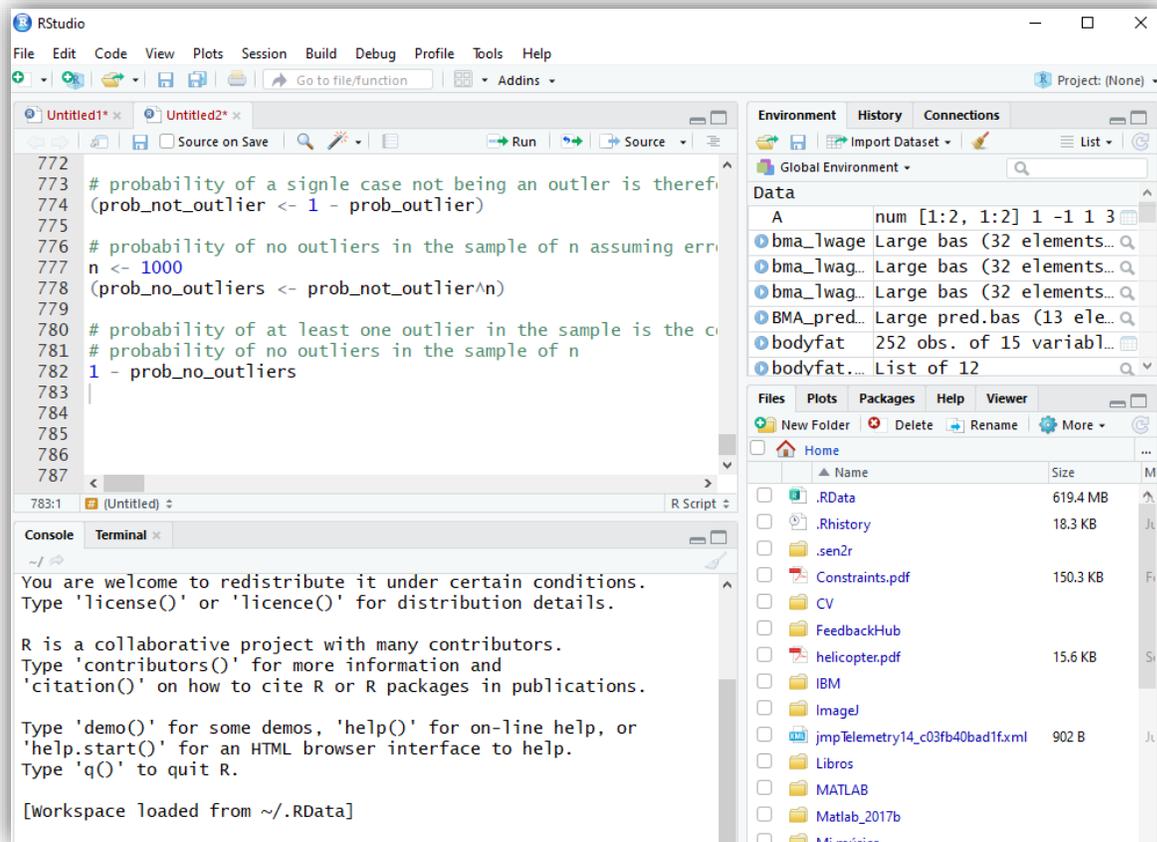
<https://www.r-project.org/>

# Por que aprender R?

---



# Por que aprender R?



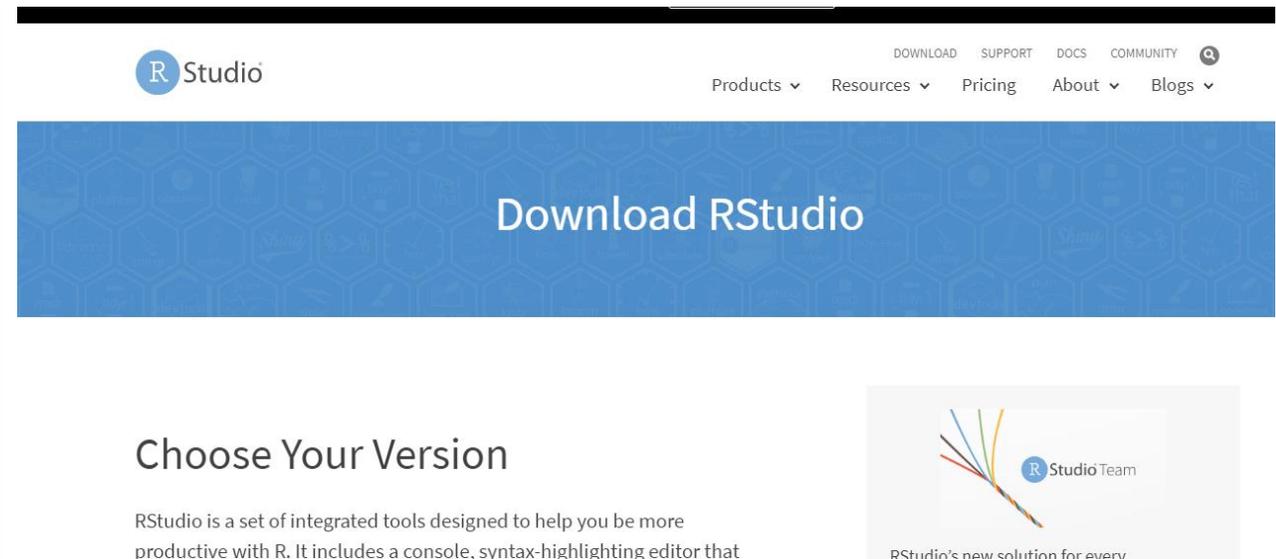
The screenshot shows the RStudio environment with a script editor, environment pane, and console. The script editor contains R code for outlier detection. The environment pane shows the Global Environment with a Data table. The console shows the R startup message.

```
772  
773 # probability of a single case not being an outlier is therefore  
774 (prob_not_outlier <- 1 - prob_outlier)  
775  
776 # probability of no outliers in the sample of n assuming error  
777 n <- 1000  
778 (prob_no_outliers <- prob_not_outlier^n)  
779  
780 # probability of at least one outlier in the sample is the complement  
781 # probability of no outliers in the sample of n  
782 1 - prob_no_outliers  
783  
784  
785  
786  
787
```

Environment: Global Environment  
Data: A num [1:2, 1:2] 1 -1 1 3  
bma\_lwage Large bas (32 elements...  
bma\_lwag... Large bas (32 elements...  
bma\_lwag... Large bas (32 elements...  
BMA\_pred... Large pred.bas (13 ele...  
bodyfat 252 obs. of 15 variabl...  
bodyfat... List of 12

Files: Home  
Name Size  
.RData 619.4 MB  
.Rhistory 18.3 KB  
.sen2r  
Constraints.pdf 150.3 KB  
CV  
FeedbackHub  
helicopter.pdf 15.6 KB  
IBM  
ImageJ  
jmpTelemetry14\_c03fb40bad1f.xml 902 B  
Libros  
MATLAB  
Matlab\_2017b  
Música

Console: You are welcome to redistribute it under certain conditions. Type 'license()' or 'licence()' for distribution details.  
R is a collaborative project with many contributors. Type 'contributors()' for more information and 'citation()' on how to cite R or R packages in publications.  
Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.  
[Workspace loaded from ~/.RData]



The screenshot shows the RStudio website homepage. The main heading is "Download RStudio". Below it, there is a section titled "Choose Your Version" with a sub-heading "RStudio's new solution for every".

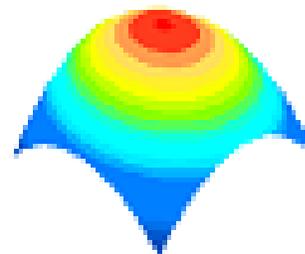
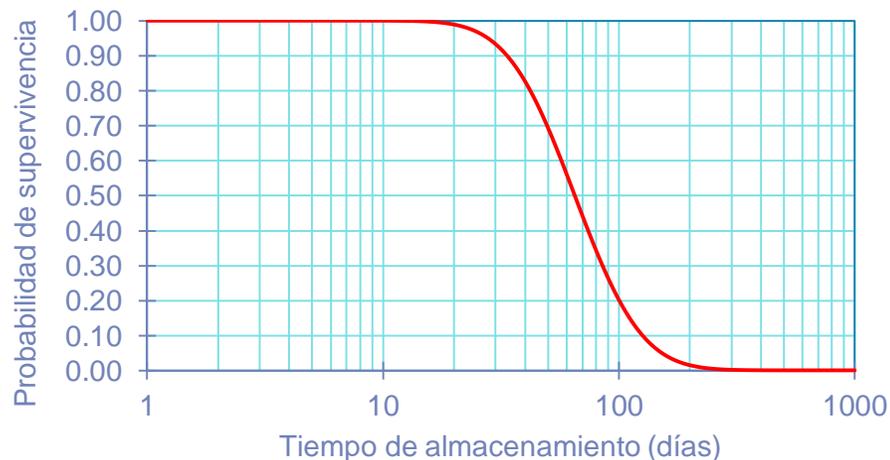
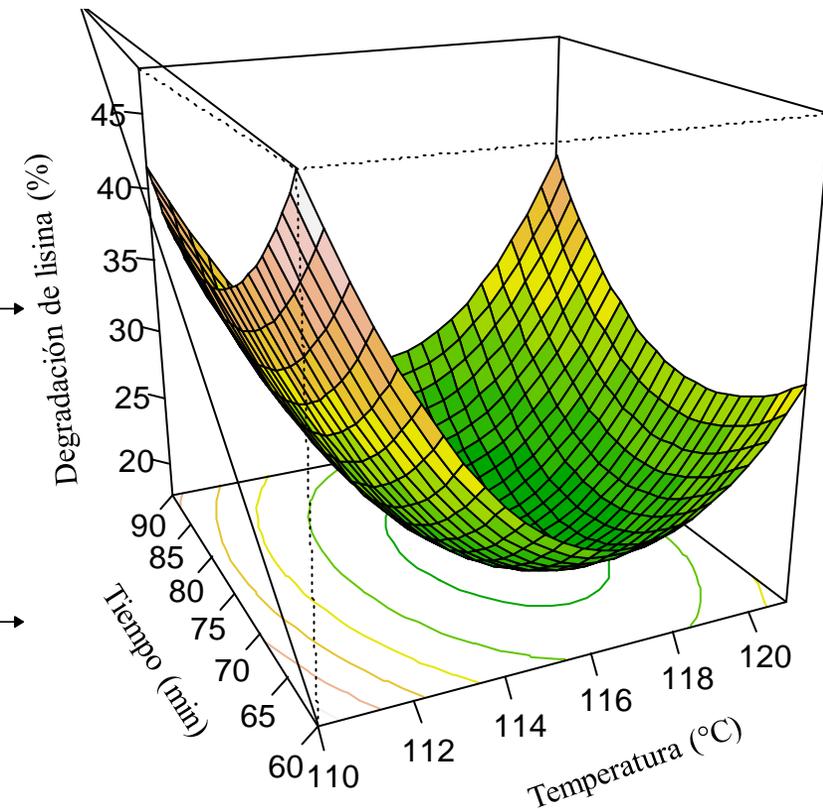
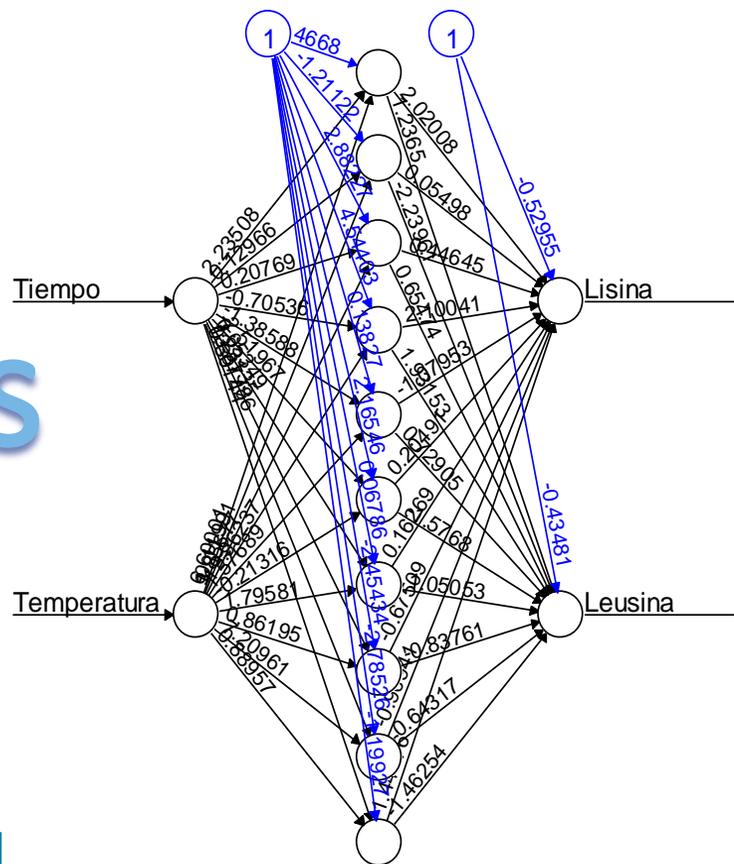
## Download RStudio

### Choose Your Version

RStudio is a set of integrated tools designed to help you be more productive with R. It includes a console, syntax-highlighting editor that

<https://rstudio.com/products/rstudio/download/>

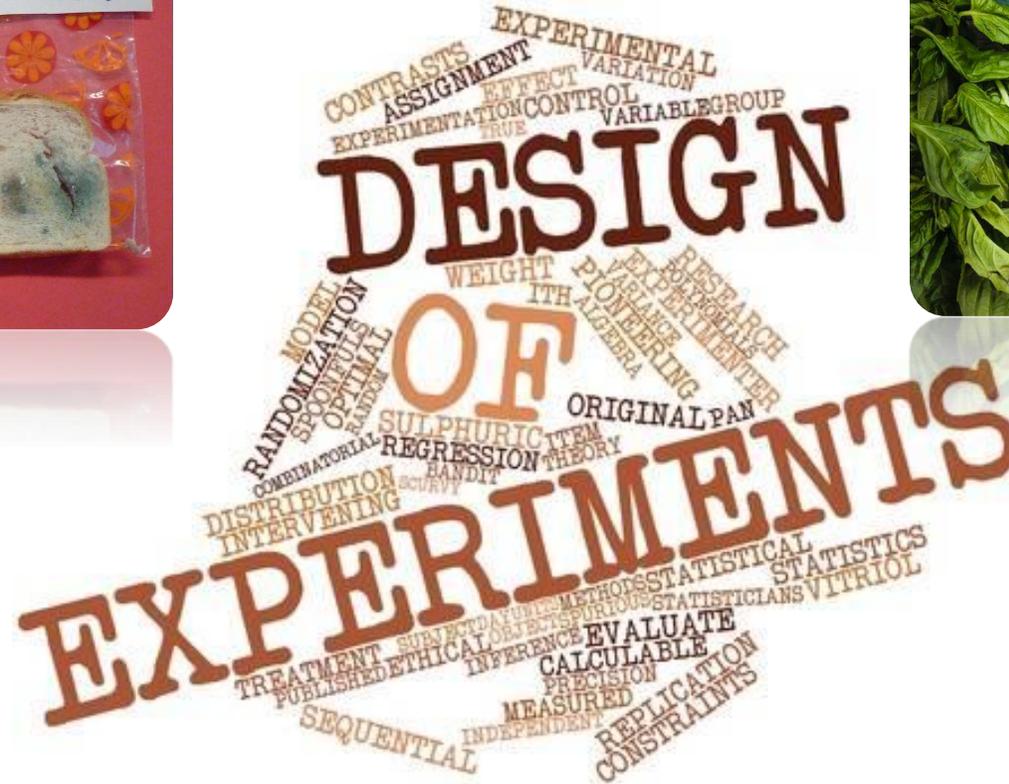
# Diseño de experimentos con



**Data Engineering**  
*Soluciones e Ingeniería*



# Temario 1. Introducción al diseño de experimentos



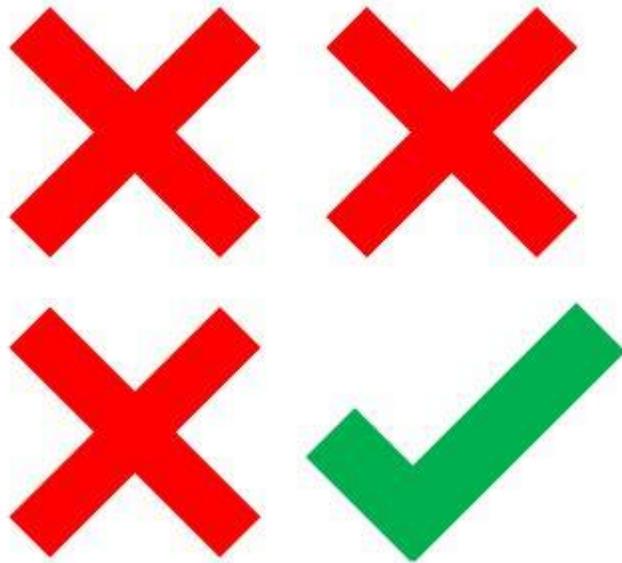
# Introducción al diseño de experimentos

Consiste en planear y realizar un conjunto de pruebas con el objetivo de generar datos que, al ser analizados estadísticamente, proporcionen evidencias objetivas que permitan responder las interrogantes planteadas por el experimentador sobre determinada situación.

1. El diseño de experimentos en la actualidad.
2. Definiciones básicas en el diseño de experimentos.
3. Etapas en el diseño de experimentos.

# 1. El diseño de experimentos en la actualidad

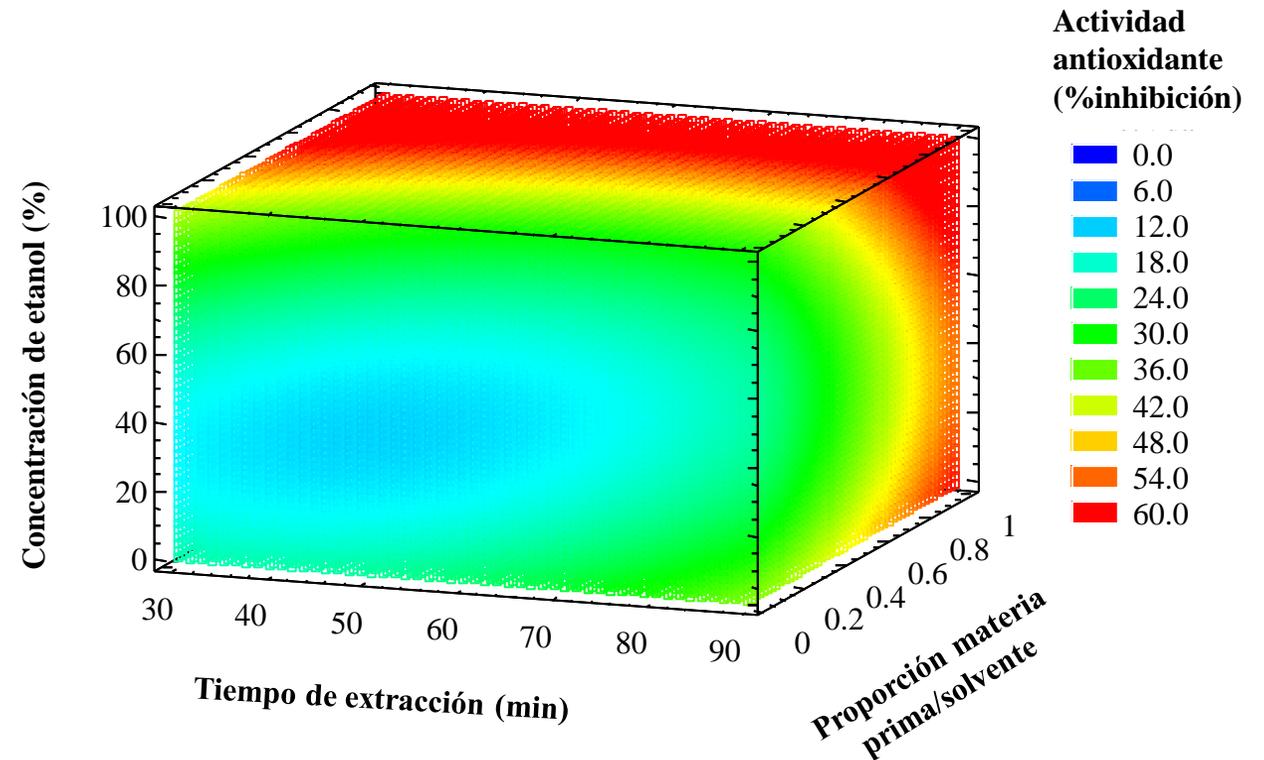
---



Es común que las pruebas o experimentos se hagan sobre la marcha, con base en el ***ensayo y error***, apelando a la experiencia y a la intuición.

# 1. El diseño de experimentos en la actualidad

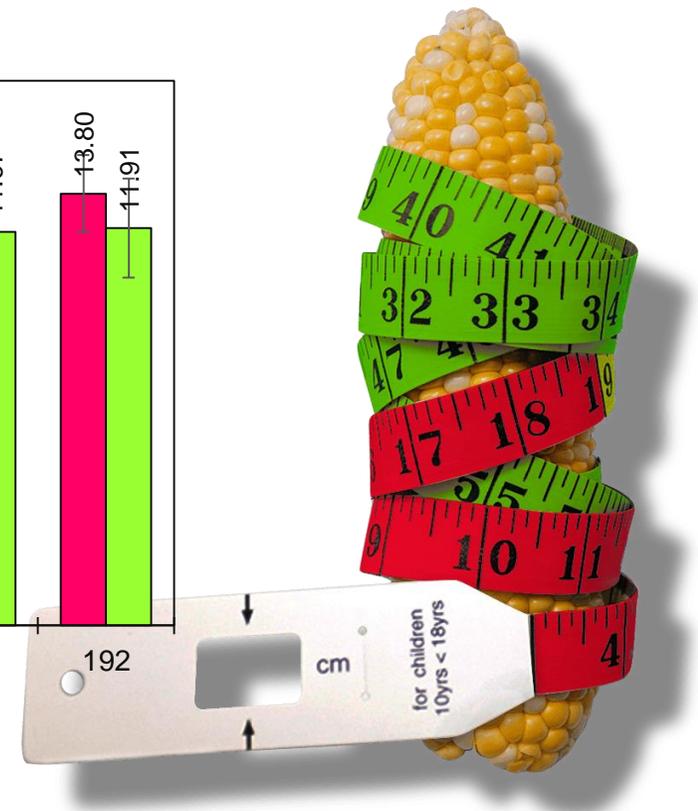
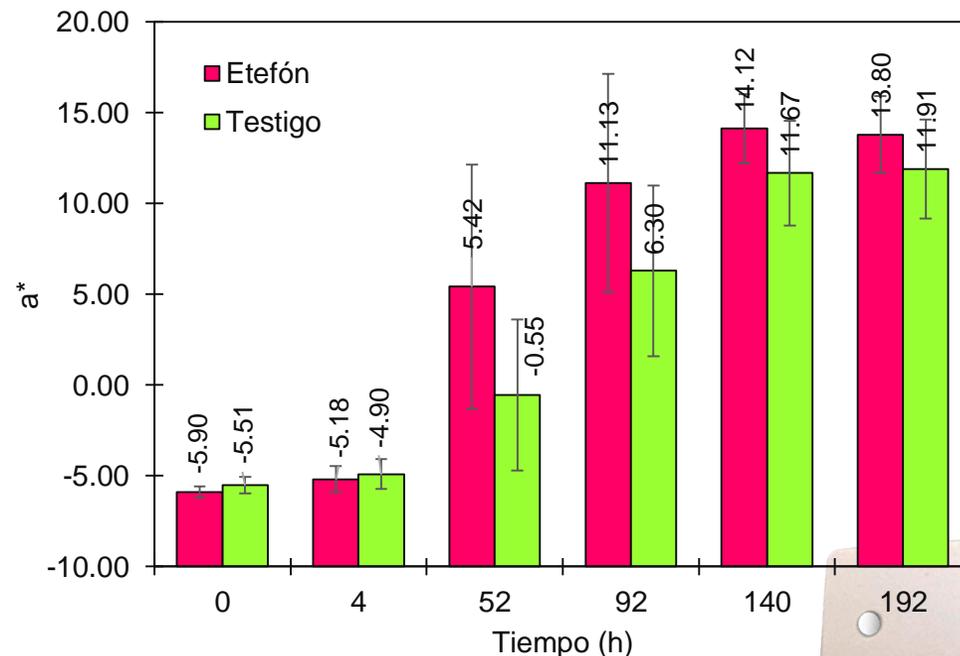
Std	Run	A:Aloe vera	B:Gelatina	C:Glicerol
6	1	15.000	0.625	0.050
16	2	10.000	0.625	0.125
14	3	10.000	0.625	0.125
1	4	5.000	0.250	0.125
9	5	10.000	0.250	0.050
2	6	15.000	0.250	0.125
12	7	10.000	1.000	0.200
13	8	10.000	0.625	0.125
17	9	10.000	0.625	0.125
5	10	5.000	0.625	0.050
7	11	5.000	0.625	0.200
3	12	5.000	1.000	0.125
10	13	10.000	1.000	0.050
11	14	10.000	0.250	0.200
8	15	15.000	0.625	0.200
4	16	15.000	1.000	0.125
15	17	10.000	0.625	0.125



El diseño de experimentos consiste en determinar cuáles pruebas se deben realizar y de qué manera, para obtener datos que, al ser analizados estadísticamente, proporcionen **evidencias objetivas** que permitan responder las interrogantes planteadas.

# 1. El diseño de experimentos en la actualidad

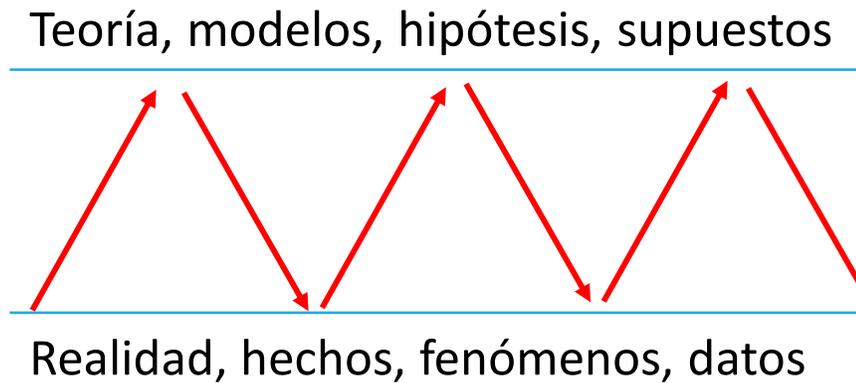
El saber diseño de experimentos y otras técnicas estadísticas, en combinación con conocimientos del proceso, sitúan al responsable del mismo como un observador perceptivo y proactivo que es capaz de proponer mejoras y de observar algo interesante (oportunidades de mejora) en el proceso y en los datos donde otra persona no ve nada.



# 1. El diseño de experimentos en la actualidad

---

## Proceso interactivo de la experimentación



# 1. El diseño de experimentos en la actualidad

## Breve historia del diseño de experimentos



Ronald A. Fisher

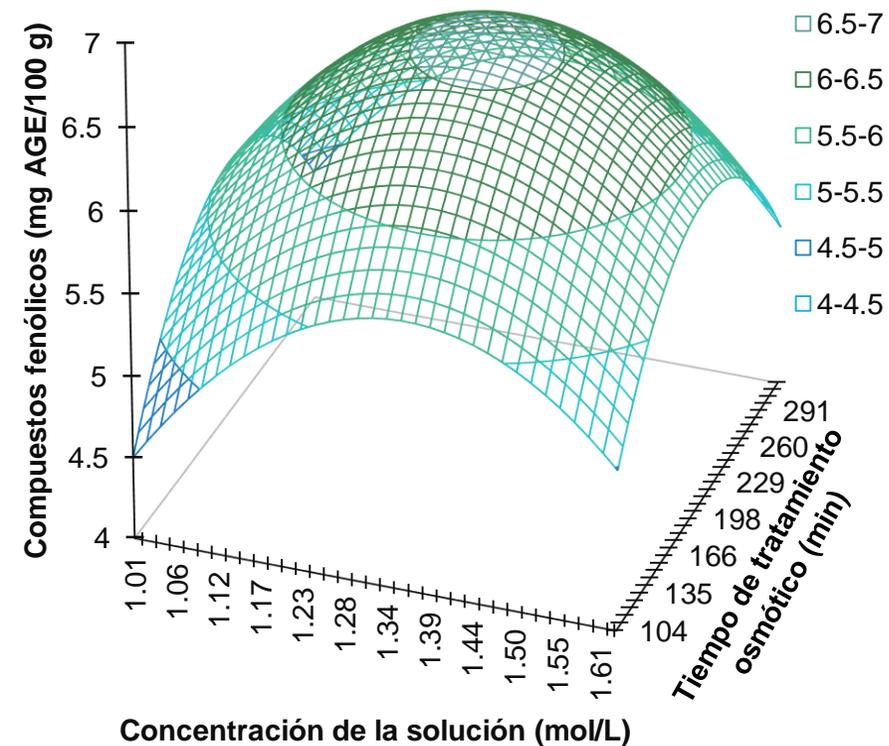


# 1. El diseño de experimentos en la actualidad

## Breve historia del diseño de experimentos



George E. P. Box

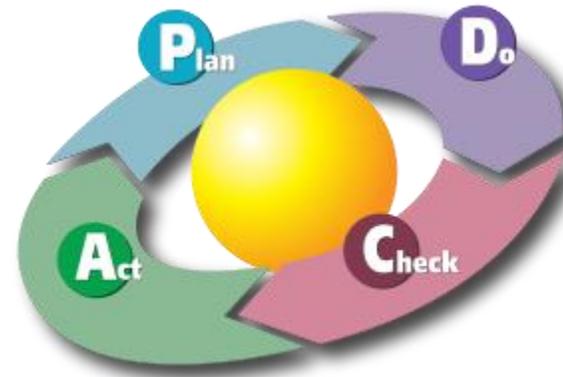


# 1. El diseño de experimentos en la actualidad

## Breve historia del diseño de experimentos



William Edwards Deming

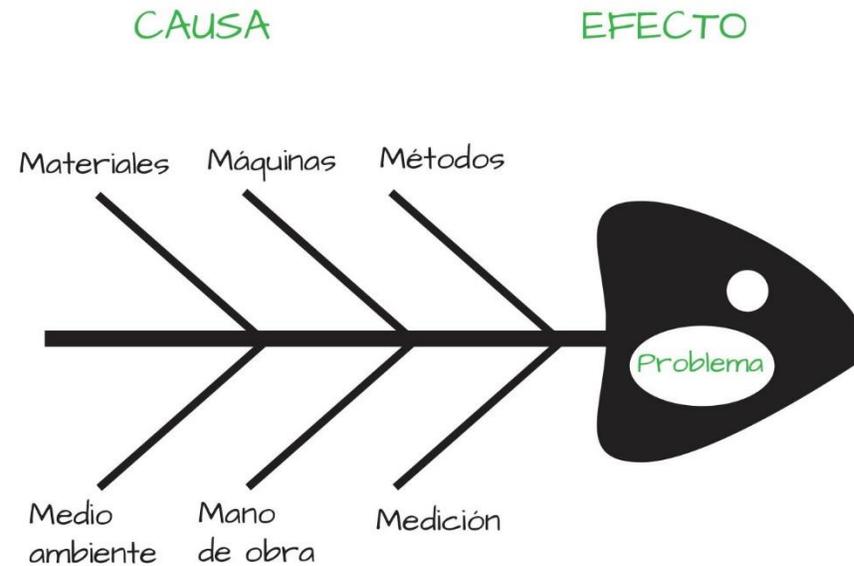


# 1. El diseño de experimentos en la actualidad

## Breve historia del diseño de experimentos



Kaoru Ishikawa

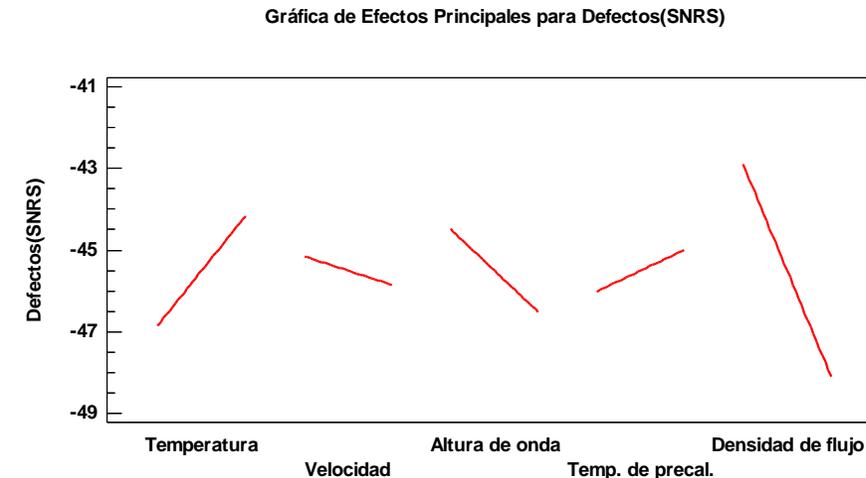


# 1. El diseño de experimentos en la actualidad

## Breve historia del diseño de experimentos



Genichi Taguchi



***“La falta de calidad de un producto es la pérdida que su utilización causa en la sociedad”.***

***“Que es más barato rediseñar los procesos de fabricación y productos que realizar acciones de mejora de la calidad tras una auditoría”.***

## 2. Definiciones básicas en el diseño de experimentos

---



El diseño de experimentos es la aplicación del método científico para generar conocimiento acerca de un sistema o proceso, por medio de pruebas planeadas adecuadamente. Esta metodología se ha ido consolidando como un conjunto de técnicas estadísticas y de ingeniería, que permiten entender mejor situaciones complejas de relación causa-efecto.

## 2. Definiciones básicas en el diseño de experimentos

---

### Experimento

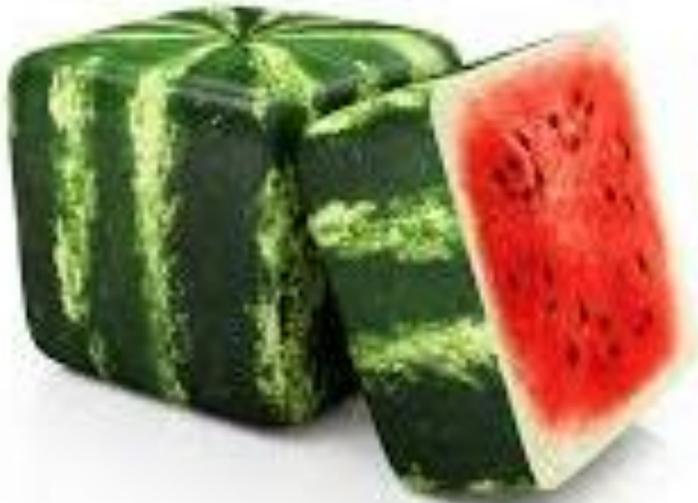
Es un cambio en las condiciones de operación de un sistema o proceso, que se hace con el objetivo de medir el efecto del cambio en una o varias propiedades del producto o resultado.



## 2. Definiciones básicas en el diseño de experimentos

---

### Experimento



## 2. Definiciones básicas en el diseño de experimentos

---

### Unidad experimental

Pieza(s) o muestra(s) que se utiliza para generar un valor que sea representativo del resultado de la prueba.



## 2. Definiciones básicas en el diseño de experimentos

### Variables, factores y niveles

#### Entrada

Factores controlables →

Factores no controlables →

Causas →



#### Salida

→ Características de calidad o variables de respuesta

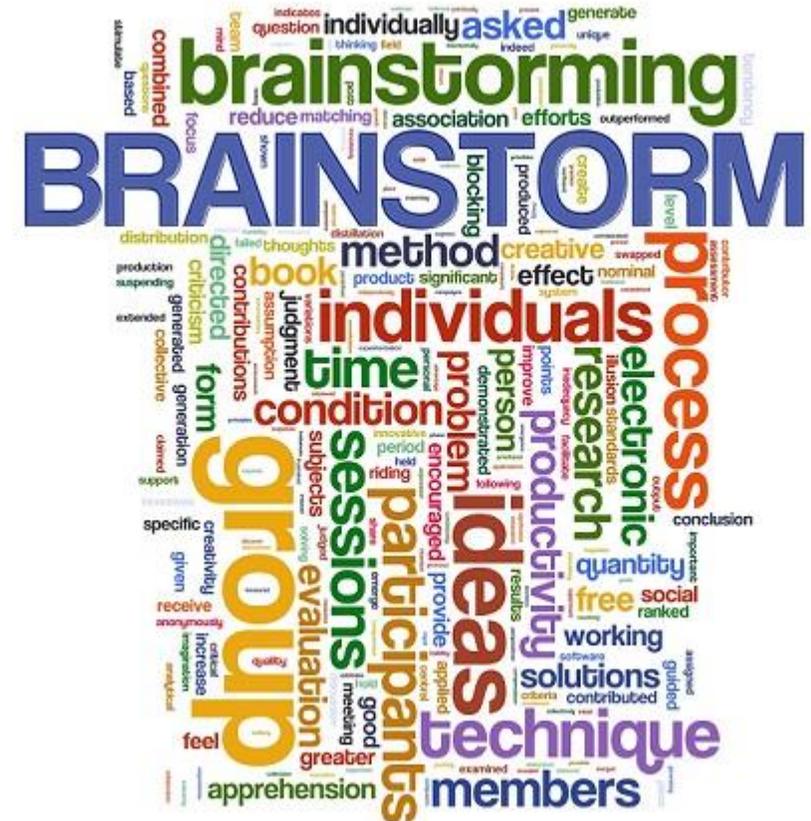
→ Efectos

- ¿Cuáles características de calidad se van a medir?
- ¿Cuáles factores controlables deben incluirse en el experimento?
- ¿Qué niveles debe utilizar cada factor?
- ¿Cuál diseño experimental es el adecuado?

# 3. Etapas en el diseño de experimentos

## Planeación y realización

Entender y delimitar el problema u objeto de estudio

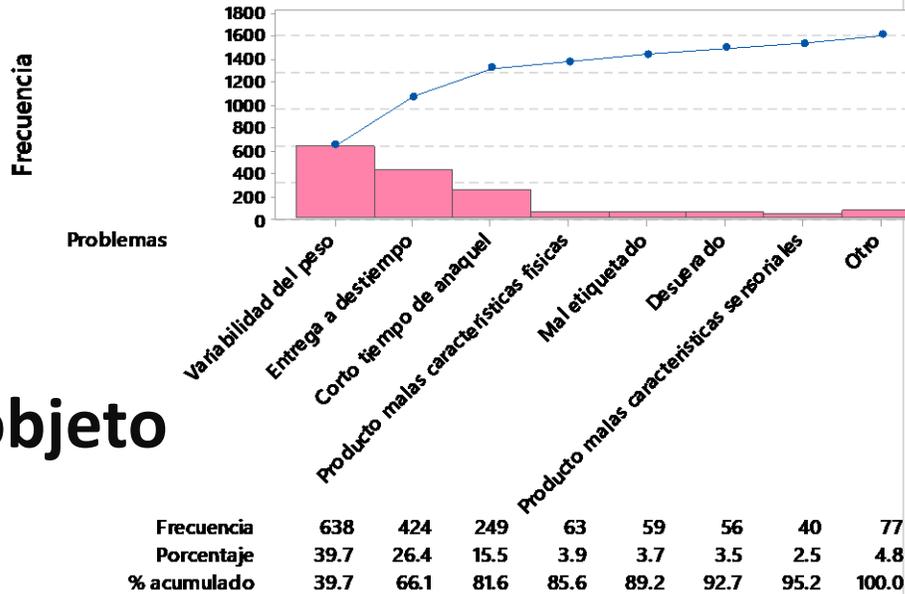


# 3. Etapas en el diseño de experimentos

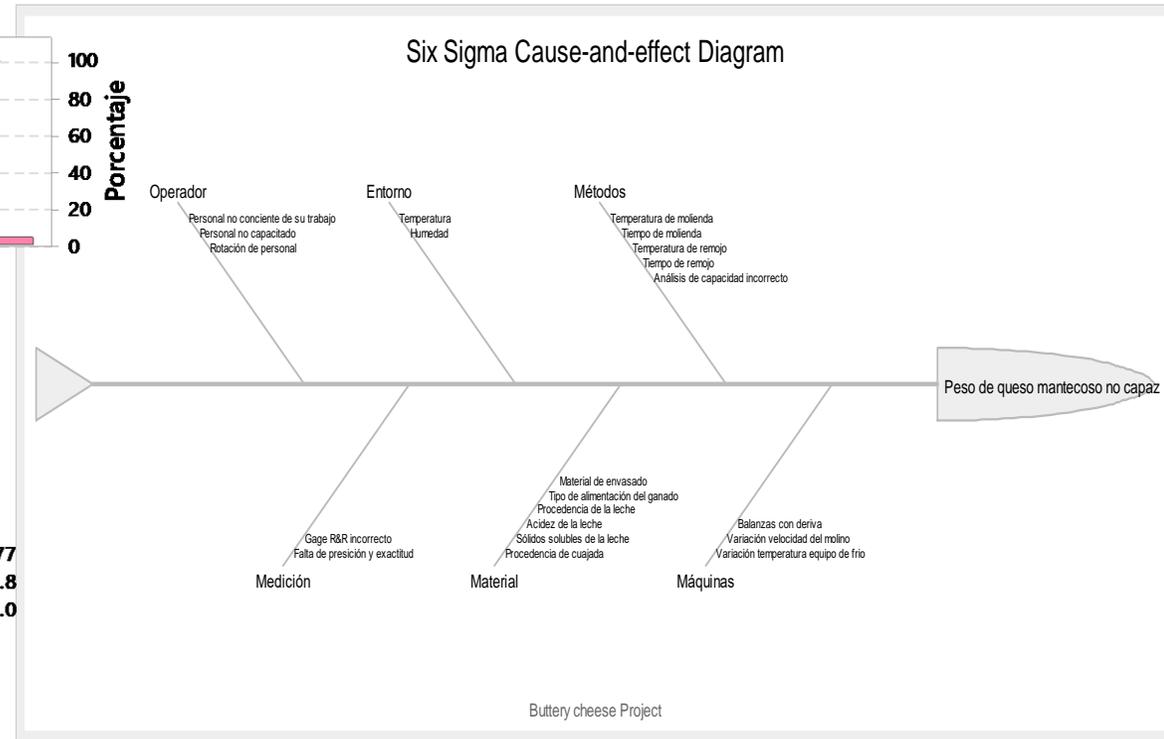
## Planeación y realización

Entender y delimitar el problema u objeto de estudio

Diagrama de Pareto de Problemas



Six Sigma Cause-and-effect Diagram



# 3. Etapas en el diseño de experimentos

---

## Planeación y realización

**Determinar cuáles factores deben estudiarse o investigarse, de acuerdo a la supuesta influencia que tienen sobre la respuesta.**



# 3. Etapas en el diseño de experimentos

## Planeación y realización

Seleccionar los niveles de cada factor, así como el diseño experimental adecuado a los factores que se tienen y al objetivo del experimento

Nivel de velocidad (RPM)	Nivel de temperatura (°C)
80	60
160	60
80	90
160	90

Nivel de velocidad (RPM)	Nivel de temperatura (°C)
80	60
160	60
80	90
160	90

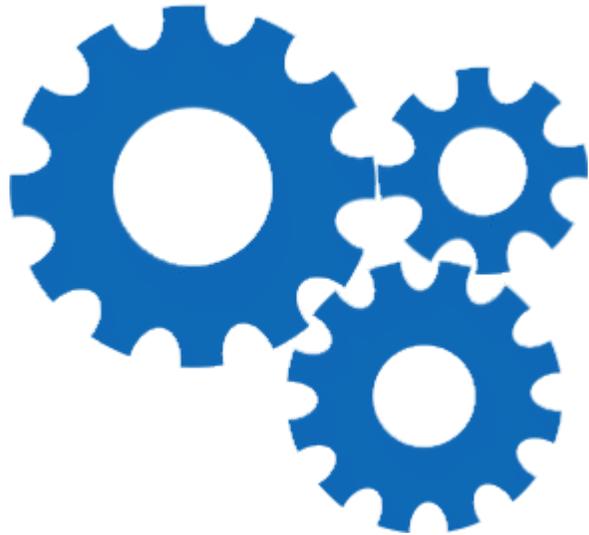
Nivel de velocidad (RPM)	Nivel de temperatura (°C)
80	60
160	60
80	90
160	90

# 3. Etapas en el diseño de experimentos

---

## Planeación y realización

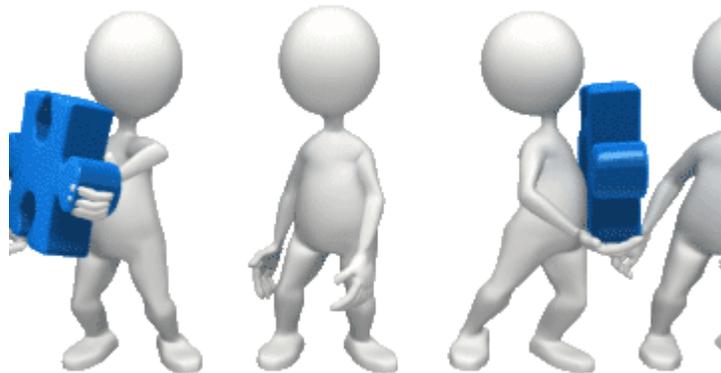
Realizar el experimento



# 3. Etapas en el diseño de experimentos

## Planeación y realización

Planear y organizar el trabajo experimental



Trat	Concentración de la solución (mol/L)		Tiempo de tratamiento osmótico (min)	
	Codificado	Real	Codificado	Real
1	-1	1.01	-1	103.60
2	1	1.63	-1	103.60
3	-1	1.01	1	316.40
4	1	1.63	1	316.40
5	-1.414	0.88	0	210.00
6	1.414	1.76	0	210.00
7	0	1.32	-1.414	59.53
8	0	1.32	1.414	360.47
9	0	1.32	0	210.00
10	0	1.32	0	210.00
11	0	1.32	0	210.00
12	0	1.32	0	210.00



# 3. Etapas en el diseño de experimentos

## Análisis

Origen	Grados de libertad	Suma de cuadrados	Media cuadrática	F	p	R <sup>2</sup>	R <sup>2</sup> -ajus
Regresión	5	15877050.182	3175410.036	6.670	0.019		
Residuos	6	2856504.584	476084.097			0.85	0.72
Total	11	18733554.765					

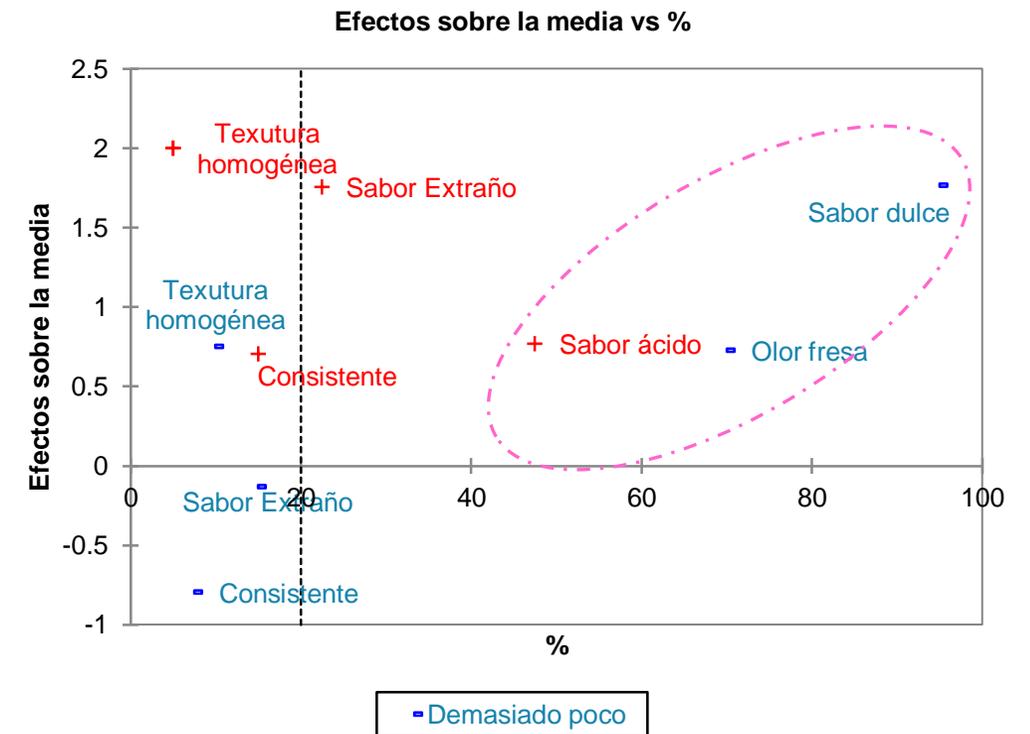
Origen	Suma de cuadrados	Grados de libertad	Media cuadrática	F	p
pH	962.197	1	962.197	370.148	0.000
Goma	3254.205	2	1627.103	625.931	0.000
pH * Goma	1837.421	2	918.710	353.419	0.000
Error	31.194	12	2.599		
Total	6085.017	17			

Fuente	Suma de cuadrados	Grados de libertad	Media cuadrática	F	p	R <sup>2</sup>	R <sup>2</sup> -ajus
Media	1606000.0	1	1606000.0				
Lineal	35914.2	2	17957.1	1.090	0.386	0.238	0.020
Cuadrático	105900.0	3	35292.6	15.550	0.011	0.940	0.865
Cúbico especial	1875.2	1	1875.2	0.780	0.442	0.952	0.857
Cúbico completo	3768.3	2	1884.1	0.550	0.691	0.977	0.795
Residual	3437.2	1	3437.2				
Total	1757000.0	10	175700.0				

# 3. Etapas en el diseño de experimentos

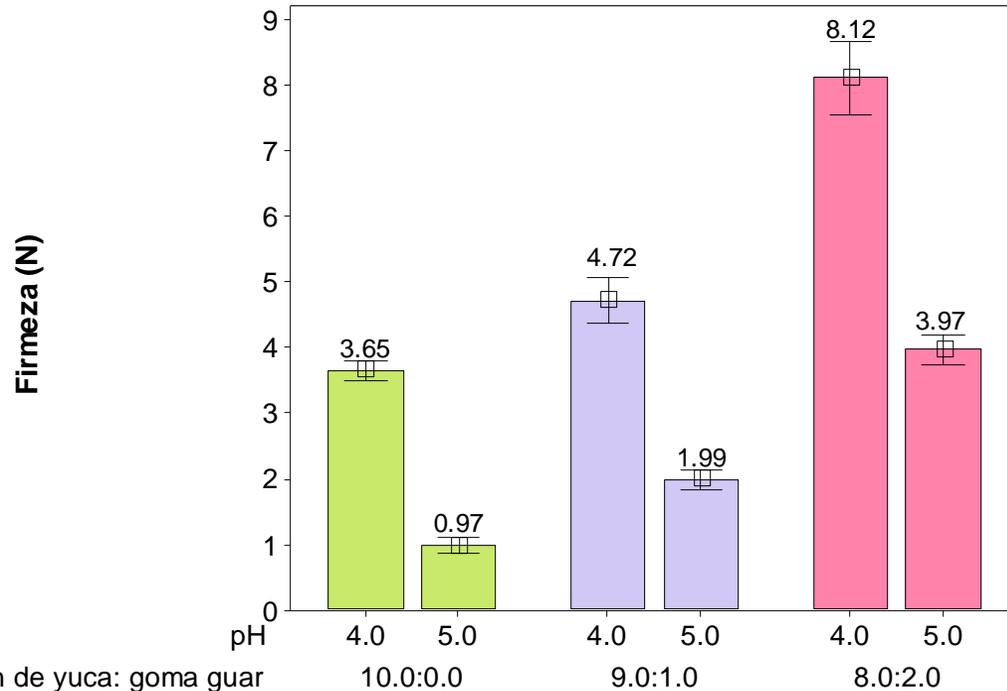
## Interpretación

Variable	Nivel	%	Suma (Aceptabilidad )	Media (Aceptabilidad )	Efectos sobre la media	p
Textura homogénea	Demasiado poco	10.00%	21.0000	5.2500	0.7500	
	JAR	85.00%	204.0000	6.0000		
	Demasiado	5.00%	8.0000	4.0000	2.0000	
Olor fresa	Demasiado poco	70.00%	157.0000	5.6071	0.7262	0.153
	JAR	30.00%	76.0000	6.3333		
Sabor dulce	Demasiado poco	95.00%	218.0000	5.7368	1.7632	0.095
	JAR	5.00%	15.0000	7.5000		
	Demasiado	0.00%				
Consistente	Demasiado poco	7.50%	20.0000	6.6667	-0.7957	
	JAR	77.50%	182.0000	5.8710		
	Demasiado	15.00%	31.0000	5.1667	0.7043	
Sabor ácido	Demasiado poco	0.00%				
	JAR	52.50%	130.0000	6.1905		
	Demasiado	47.50%	103.0000	5.4211	0.7694	0.098
Sabor Extraño	Demasiado poco	15.00%	38.0000	6.3333	-0.1333	
	JAR	62.50%	155.0000	6.2000		
	Demasiado	22.50%	40.0000	4.4444	1.7556	0.000



# 3. Etapas en el diseño de experimentos

## Interpretación



pH	Proporción almidón de yuca: goma guar	Subgrupo				
		1	2	3	4	5
5.0	10.0:0.0	0.97				
5.0	9.0:1.0		1.99			
4.0	10.0:0.0			3.65		
5.0	8.0:2.0			3.97	3.97	
4.0	9.0:1.0				4.72	
4.0	8.0:2.0					8.12

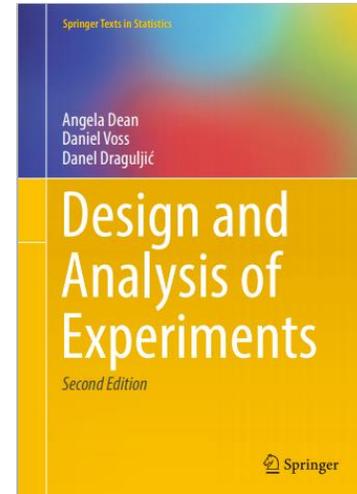
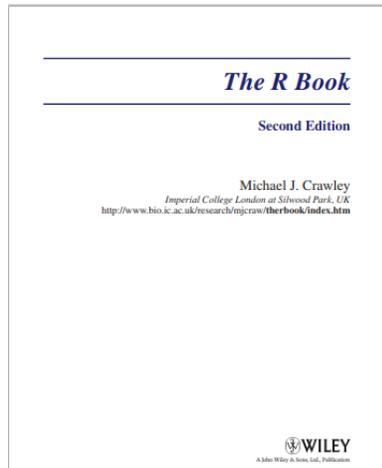
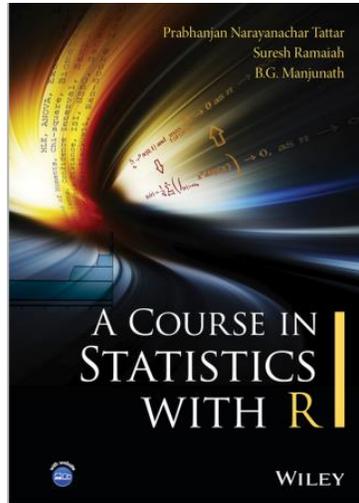
# 3. Etapas en el diseño de experimentos

---

## Control y conclusiones finales



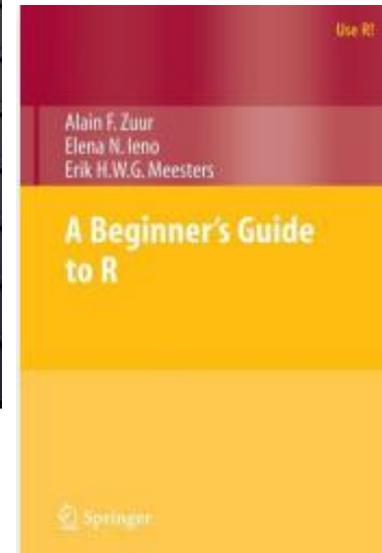
# Donde aprender diseños de experimento y R?



R Programming  
for Data Science



Roger D. Peng

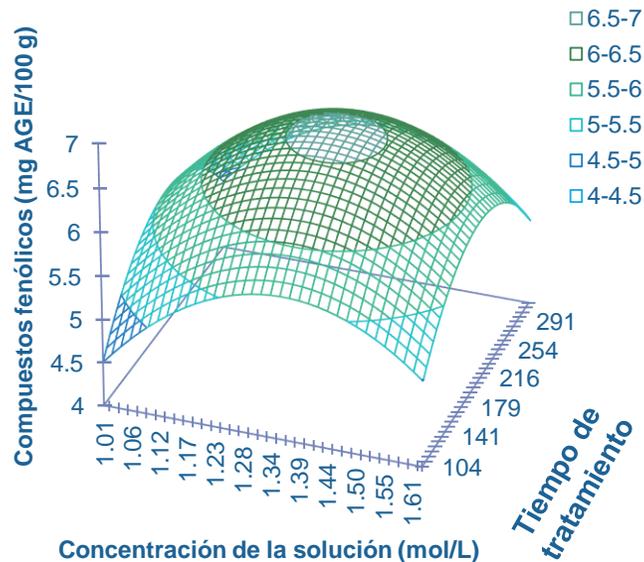


**course**ra

**edX**

Taller online gratuito:

# Diseño de experimentos con R



## Expositores:



**Ing. Mg. Jesús Alfredo Obregón Domínguez**

- Ingeniero en Industrias Alimentarias.
- Maestro en Ciencias con mención es Estadística Aplicada - Universidad Nacional de Trujillo.
- Gerente General de Data Engineering.
- Especialista en diseño de experimentos.

**Ing. Jhon Aguilar Castillo**

- Ing. Estadístico-UNT.
- Maestría en Informática-PUCP.
- Especialista estadístico-MINEDU.
- Experiencia en Docencia estadística e investigación-UNTRM.

Conferencia vía:



PONTIFICIA  
UNIVERSIDAD  
CATÓLICA  
DEL PERÚ

**Fecha: martes 7 de julio de 20:00 a 21:30.  
sábado 11 de julio de 15:00 a 16:30.**