



El Diseño de experimentos como herramienta óptima en la investigación.

Profesor: Xavier Barber Vallés

I.U.I. Centro de Investigación Operativa

Universidad Miguel Hernández de Elche

Valencia Bayesian Research Group

<http://vabar.github.io/>



Índice

- Planificación de la Investigación.
- Los datos.
- Analizando los datos de forma clásica.
- Más allá del ANOVA, la no linealidad.
- El diseño de experimentos como problema de decisión.



<https://creativecommons.org/2014/11/26/cc-goes-to-mozfest-2014/>

Planificación de la Investigación

Qué, Cómo, Dónde y Cuánto planificar

Primera Pregunta

¿Empezamos la casa por el tejado?

¿O la empezamos por unos BUENOS cimientos?



<https://pixabay.com/es/casa-rural-casa-al-rev%C3%A9s-928979/>

Segunda Pregunta

¿Existe
disponibilidad
económica par
ala realización
del estudio?



<http://pixabay.com/es/hombre-la-figura-juguetes-76196/>



Impossible is nothing

<http://theninjutsuartist.deviantart.com/art/Impossible-is-Nothing-289733564>

PUES ... ES POSIBLE



Planificando

- ¿Qué queremos hacer?
- ¿Por qué lo queremos hacer?
- ¿Cómo lo queremos hacer?
- ¿Cuándo lo vamos a hacer?
- ¿Dónde lo vamos a hacer?

¿Qué queremos hacer?

Nuestra “pregunta” de investigación o desarrollo de producto debe de ir siempre acompañada de una base científica.

Es muy importante, tener claro el fin.

¿Por qué queremos hacerlo?

¿Alguien en otro lugar del planeta se ha planteado esta mismo “qué”?

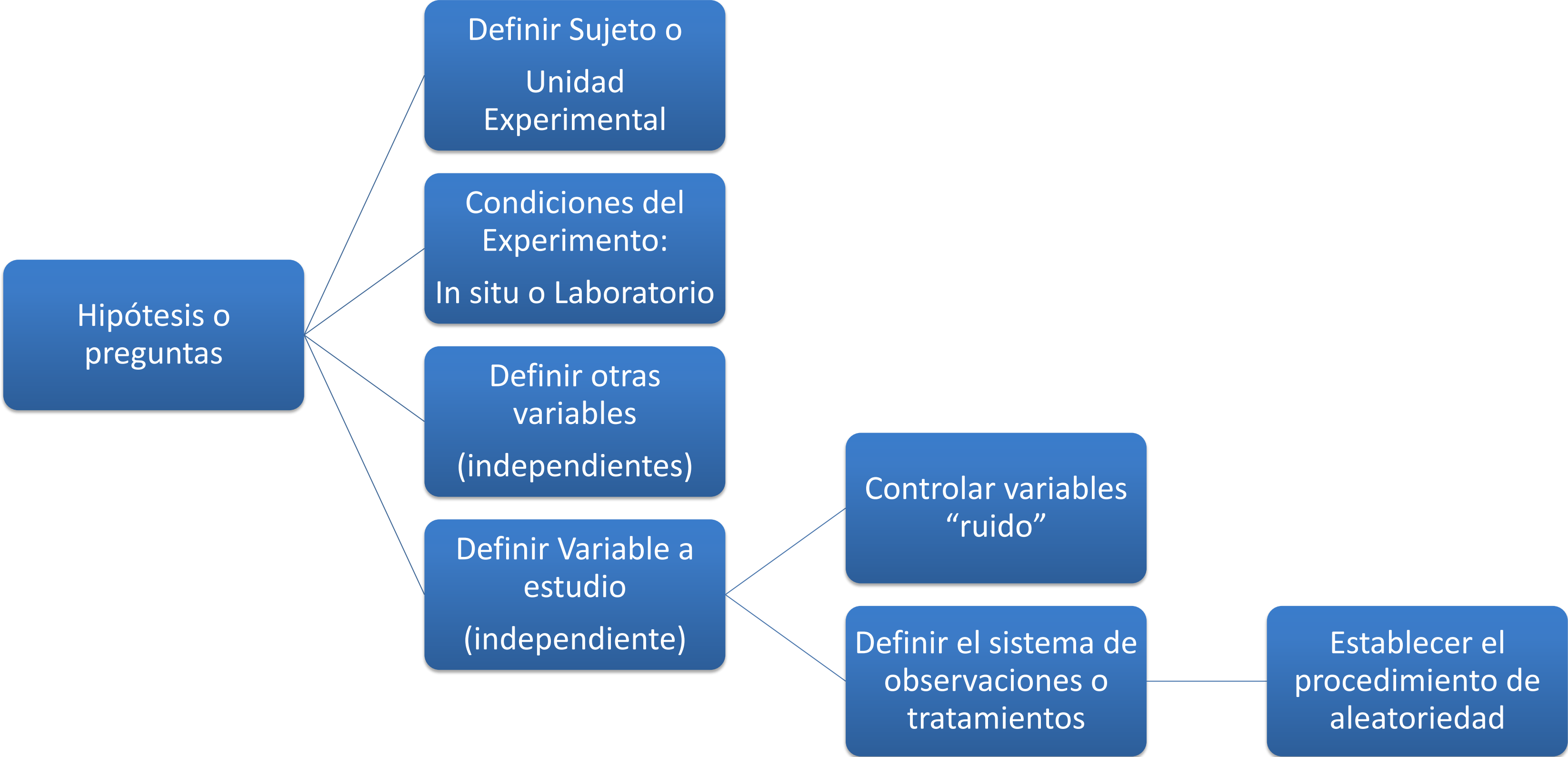
- Búsqueda bibliográfica **EXHAUSTIVA**
 - Buscadores de internet
 - Revistas Científicas
 - Revistas de Divulgación
 - Oficinas de Patentes
 - Etc.

¿Cómo?

Este es el punto más importante si se desea tener un resultado reproducible.

- **Reproducibilidad**
- **“Potencia” de las conclusiones**

A raíz de la publicación de un artículo en *Nature Reviews Neuroscience* en el que se cuestiona la validez de las conclusiones obtenidas en el ámbito de las neurociencias (Button et al., 2013). [Sólo un 21% válidos!!!!]



¿Cómo?

Utilizando el legado de R.A. Fisher:

- En octubre de 1919, Fisher fue contratado en Rothamsted Experimental Station.
- Le pidieron que trabajara con ellos de seis meses a un año (datos agrícolas).
- En 1926 publicó el primer resumen completo de sus ideas en el artículo “The Arrangement of Field Experiments”, donde describió control total , repetición y aleatorización.

...y la estadística del siglo XXI

¿Cómo?

- **Documentar:**
 - Objetivos específicos del experimento,
 - identificación de los factores que influyen y cuáles de ellos varían y cuáles permanecen constantes,
 - características a medir,
 - procedimientos particulares para realizar las pruebas o medir las características,
 - número de repeticiones del experimento básico a realizar, recursos y materiales disponibles.

¿Cómo?

Experimento:

- conjunto de circunstancias creados para el experimento, en respuesta a la hipótesis de investigación y son el centro de la misma.

Tratamientos:

- la entidad física o sujeto expuesto al tratamiento independientemente de otras unidades.

Unidades Experimentales:

- describe la variación entre las unidades experimentales tratadas de manera idéntica e independiente.

¿Cómo?

Control local de errores experimentales

- La técnica afecta la variación y el sesgo (error en las conclusiones).
- Selección de unidades experimentales uniformes.
- La segregación en bloques para reducir la variación del error experimental.
- El diseño experimental asigna los tratamientos.
- Añadir información adicional para el control estadístico de la variación.

¿Dónde?

En un lugar donde podamos controlar la “aleatoriedad” (que no influyan factores externos) del experimento.

Y por supuesto REPLICAR para CONTROLAR.

Recomiendo la lectura del libro: Biostatistical Design & analysis Using R, Murray Logan . Capítulo 6 , pag 134-142.



Los datos después de los “experimentos”

¿Qué hacer con los datos de las hojas Excel©?

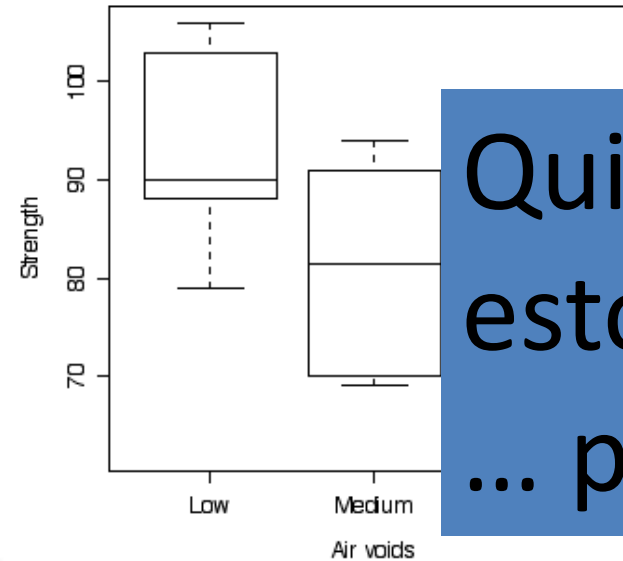
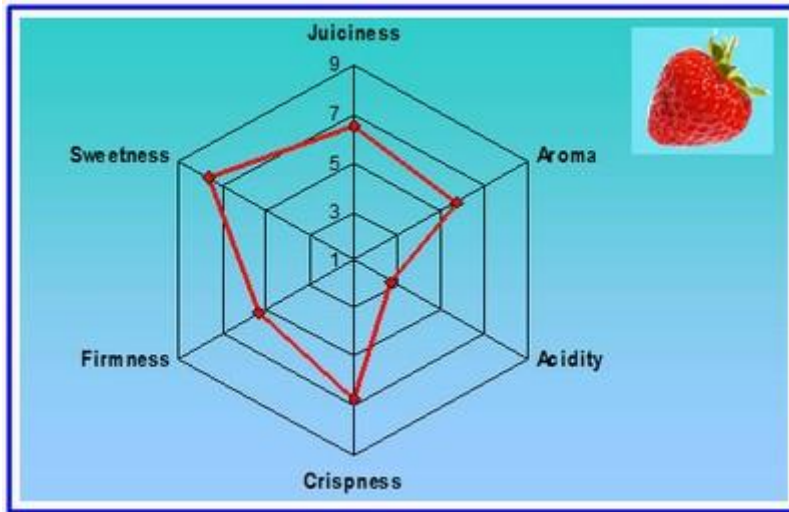
Tercera Pregunta

N°	DESCRIPTION DES OUVRAGES	Unité	Qté	Coût des fournisseurs (F.O.)		Temps M.O. (h)		Prix de revient		Prix de vente	
				Unitaire	Totale	Unit.	Totale	Main d'œuvre	M.O. + F.O.	Unitaire	Total
	- marque : FRANCE AIR, CINERGIE au 6 équivalent			Diffuseur 3 parties et 1 dar orolake							
	- 600 x 300 - M250 700x3/4 - 1 direction	U	1	125.92	125.92	1.2h	1.20h	28.80	106.87	134.00	134.00
	- 600 x 600 - M250 1500x3/4 - 1 direction	U	1	197.10	197.10	1.5h	1.50h	36.00	158.20	198.00	198.00
	Diffuseur ALU multidirectionnel, ailettes orient. + registre + pl6			tarif 35 - FRANCE AIR - GAP 31 - PFUE1 usf							
	- marque : FRANCE AIR, CINERGIE au 6 équivalent			Diffuseur 3 parties et 1 dar orolake							
	- 600 x 300 - M250 700x3/4 - 2 directions	U	1	174.24	174.24	1.2h	1.20h	28.80	136.33	171.00	171.00
	- 600 x 600 - M250 1500x3/4 - 1 direction	U	1	262.20	262.20	1.5h	1.50h	36.00	198.56	248.00	248.00
	Diffuseur ALU multidirectionnel, ailettes orient. + registre + pl6			tarif 35 - FRANCE AIR - GAP 31 - PFUE1 usf							
	- marque : FRANCE AIR, CINERGIE au 6 équivalent			Diffuseur 3 parties et 1 dar orolake							
	- 600 x 300 - M250 700x3/4 - 2 directions	U	1	128.51	128.51	1.2h	1.20h	28.80	108.48	136.00	136.00
	- 600 x 600 - M250 1500x3/4 - 2 directions	U	1	197.56	197.56	1.5h	1.50h	36.00	158.49	198.00	198.00
	Diffuseur ALU multidirectionnel, ailettes orient. + registre + pl6			tarif 35 - FRANCE AIR - GAP 31 - PFUE1 usf							
	- marque : FRANCE AIR, CINERGIE au 6 équivalent			Diffuseur 3 parties et 1 dar orolake							
	- 600 x 300 - M250 700x3/4 - 2 directions	U	1	176.83	176.83	1.2h	1.20h	28.80	138.43	173.00	173.00
	- 600 x 600 - M250 1500x3/4 - 2 directions	U	1	262.65	262.65	1.5h	1.50h	36.00	198.34	249.00	249.00
	Diffuseur ALU multidirectionnel, ailettes orient. + registre + pl6			tarif 35 - FRANCE AIR - GAP 31 - PFUE1 usf							
	- marque : FRANCE AIR, CINERGIE au 6 équivalent			Diffuseur 3 parties et 1 dar orolake							
	- 600 x 300 - M250 700x3/4 - 3 directions	U	1	110.82	110.82	1.2h	1.20h	28.80	97.51	122.00	122.00
	- 600 x 600 - M250 1500x3/4 - 3 directions	U	1	165.40	165.40	1.5h	1.50h	36.00	138.55	173.00	173.00
	Diffuseur ALU multidirectionnel, ailettes orient. + registre + pl6			tarif 35 - FRANCE AIR - GAP 31 - PFUE1 usf							
	- marque : FRANCE AIR, CINERGIE au 6 équivalent			Diffuseur 3 parties et 1 dar orolake							
	- 600 x 300 - M250 700x3/4 - 3 directions	U	1	159.15	159.15	1.2h	1.20h	28.80	127.47	159.00	159.00
	- 600 x 600 - M250 1500x3/4 - 3 directions	U	1	230.49	230.49	1.5h	1.50h	36.00	178.90	224.00	224.00
	Diffuseur ALU multidirectionnel, ailettes orient. + registre + pl6			tarif 35 - FRANCE AIR - GAP 31 - PFUE1 usf							
	- marque : FRANCE AIR, CINERGIE au 6 équivalent			Diffuseur 3 parties et 1 dar orolake							
	- 600 x 300 - M250 700x3/4 - 4 directions	U	1	112.50	112.50	1.2h	1.20h	28.80	98.55	123.00	123.00
	- 600 x 600 - M250 1500x3/4 - 4 directions	U	1	167.38	167.38	1.5h	1.50h	36.00	139.78	175.00	175.00
	Diffuseur ALU multidirectionnel, ailettes orient. + registre + pl6			tarif 35 - FRANCE AIR - GAP 31 - PFUE1 usf							
	- marque : FRANCE AIR, CINERGIE au 6 équivalent			Diffuseur 3 parties et 1 dar orolake							
	- 600 x 300 - M250 700x3/4 - 4 directions	U	1	160.82	160.82	1.2h	1.20h	28.80	128.51	161.00	161.00
	- 600 x 600 - M250 1500x3/4 - 4 directions	U	1	232.47	232.47	1.5h	1.50h	36.00	180.13	225.00	225.00
	DIFFUSEUR PLAFONNIER ARCHITECTURAUX MODULAIRE 600 x 600										
	Diffuseur blanc linéaire 600x600 + registre + pl6 num			tarif 35 - FRANCE AIR - DAP 35 - registre							
	- marque : FRANCE AIR, CINERGIE au 6 équivalent			appeler pour détail							
	- 4 directions - 1 fonte - M150 200 x 300x3/4	U	1	242.07	242.07	1.5h	1.50h	36.00	186.08	233.00	233.00
	- 4 directions - 2 fontes - M200 350 x 500x3/4	U	1	279.42	279.42	1.5h	1.50h	36.00	209.24	262.00	262.00
	- 4 directions - 3 fontes - M250 450 x 850x3/4	U	1	314.94	314.94	1.5h	1.50h	36.00	231.26	289.00	289.00
	- 4 directions - 3 fontes - M250 500 x 900x3/4	U	1	353.05	353.05	1.5h	1.50h	36.00	254.99	319.00	319.00
	Diffuseur blanc linéaire 600x600 + registre + pl6 num iral6			tarif 35 - FRANCE AIR - DAP 35 - registre usf							
	- marque : FRANCE AIR, CINERGIE au 6 équivalent			appeler pour détail							
	- 4 directions - 1 fonte - M150 200 x 300x3/4	U	1	276.53	276.53	1.5h	1.50h	36.00	207.45	259.00	259.00
	- 4 directions - 2 fontes - M200 350 x 500x3/4	U	1	322.87	322.87	1.5h	1.50h	36.00	236.18	295.00	295.00
	- 4 directions - 3 fontes - M250 450 x 850x3/4	U	1	363.57	363.57	1.5h	1.50h	36.00	261.41	327.00	327.00
	- 4 directions - 3 fontes - M250 500 x 900x3/4	U	1	400.00	400.00	1.5h	1.50h	36.00	284.00	355.00	355.00
	Diffuseur tourbillonnaire à induction, 3 diffuseurs réglable			tarif 35 - FRANCE AIR - LD 7 - tassement normal usf							
	RAL 9010 pour plafond modulaire + registre + pl6 num										
	- marque : FRANCE AIR, CINERGIE au 6 équivalent										
	- Ø125 - M... 150x3/4	U	1	123.48	123.48	1.5h	1.50h	36.00	112.56	141.00	141.00
	- Ø160 - M... 200x3/4	U	1	125.15	125.15	1.5h	1.50h	36.00	113.59	142.00	142.00

The screenshot shows Microsoft Excel with a PivotTable and the QQube Add-In interface. The PivotTable is set to show data for 'A Cheung Limited' with columns for Item Name, OpenP, Rate, and Original PO Qty. The QQube Add-In window is open, showing a list of fields to add to the report, including 'Item Name', 'Item Type', and 'Vendor Name'. A red box highlights the 'Excel Pivot Table List' in the QQube interface.

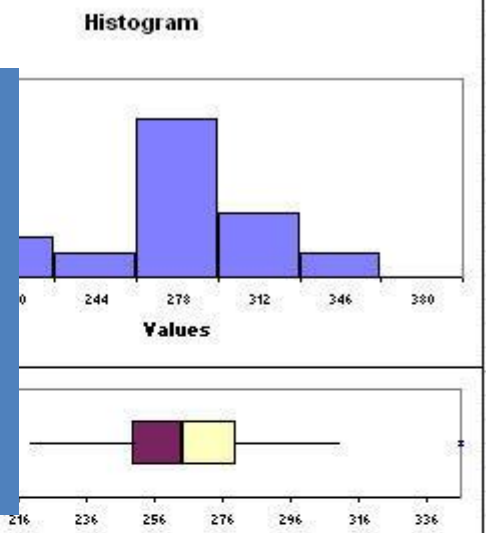
Mira que Excel® tan bonito tengo!!!!!!

Cuarta Pregunta



Quiero reproducir esto que he visto allí...
... pero con mis datos.

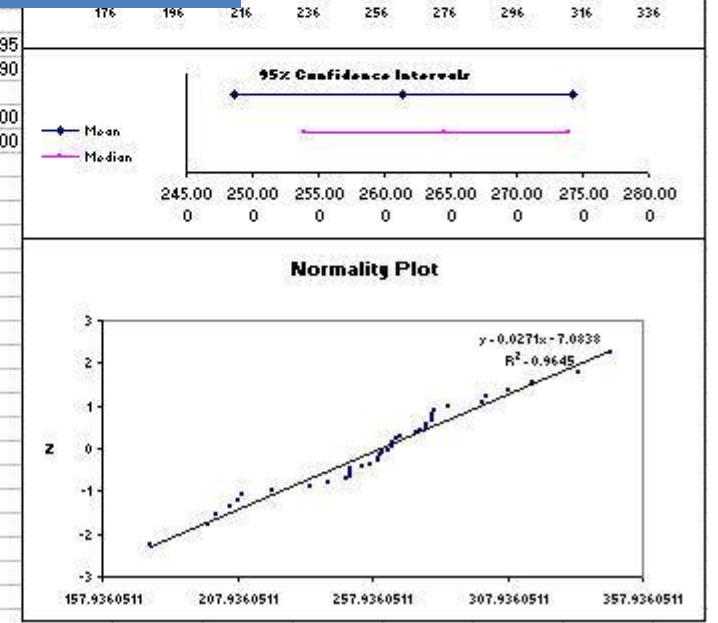
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Sample	Obs 1	Obs 2	Anderson-Darling									
2	S1	265	205	A-Squared	0.703								
3	S2	268	260	ρ	0.061								



Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Population prevalence low ac - females	64	230569	5839642	1661172	1023135.85
prevalence low ac - males	64	2.30	28.80	12.3156	6.3215
Prevalence low ac - total	64	.90	27.00	8.4594	4.7784
% measles immunization	64	43.50	100.00	80.4234	12.8637
Prev HH with iodized salt	64	.40	87.70	41.1687	21.7579
Valid N (listwise)	64				

23					
24			For Stdev (sigma)	29.595	
25				46.390	
26					
27			for Median	254.000	
28				274.000	
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					
43					
44					
45					
46					
47					
48					



Quinta pregunta



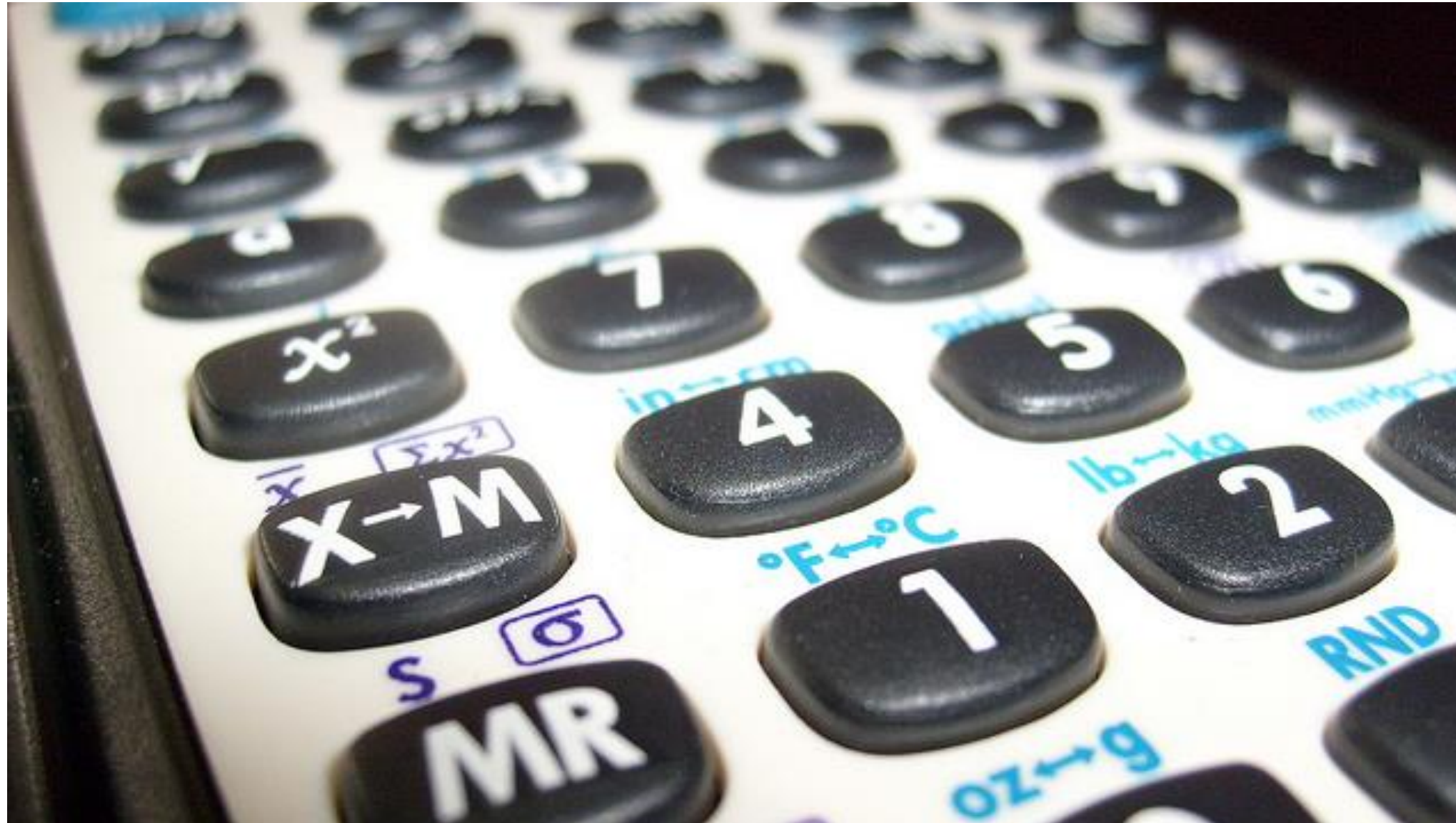
¿Eres de los que cambias la bañera por el plato de ducha tú solo?

¿Eres de los que cambias el aceite del coche tú solo?

Pues ...

¡Busca ayuda especializada!

<http://thecavernsite.blogspot.com.es/>



Analizando los datos de forma clásica.

Cuando la norma se convierte en error

ANOVA (Analysis of Variance)

- Completamente Aleatorizado
- Diseños factoriales
- Diseños factoriales: modelos aleatorios y mixtos
- Diseños por bloques: Completos e Incompletos
- Diseños factoriales fraccionarios
- Diseños con mediciones repetidas
- Diseños cruzados
- Análisis de la Covarianza

Sexta Pregunta

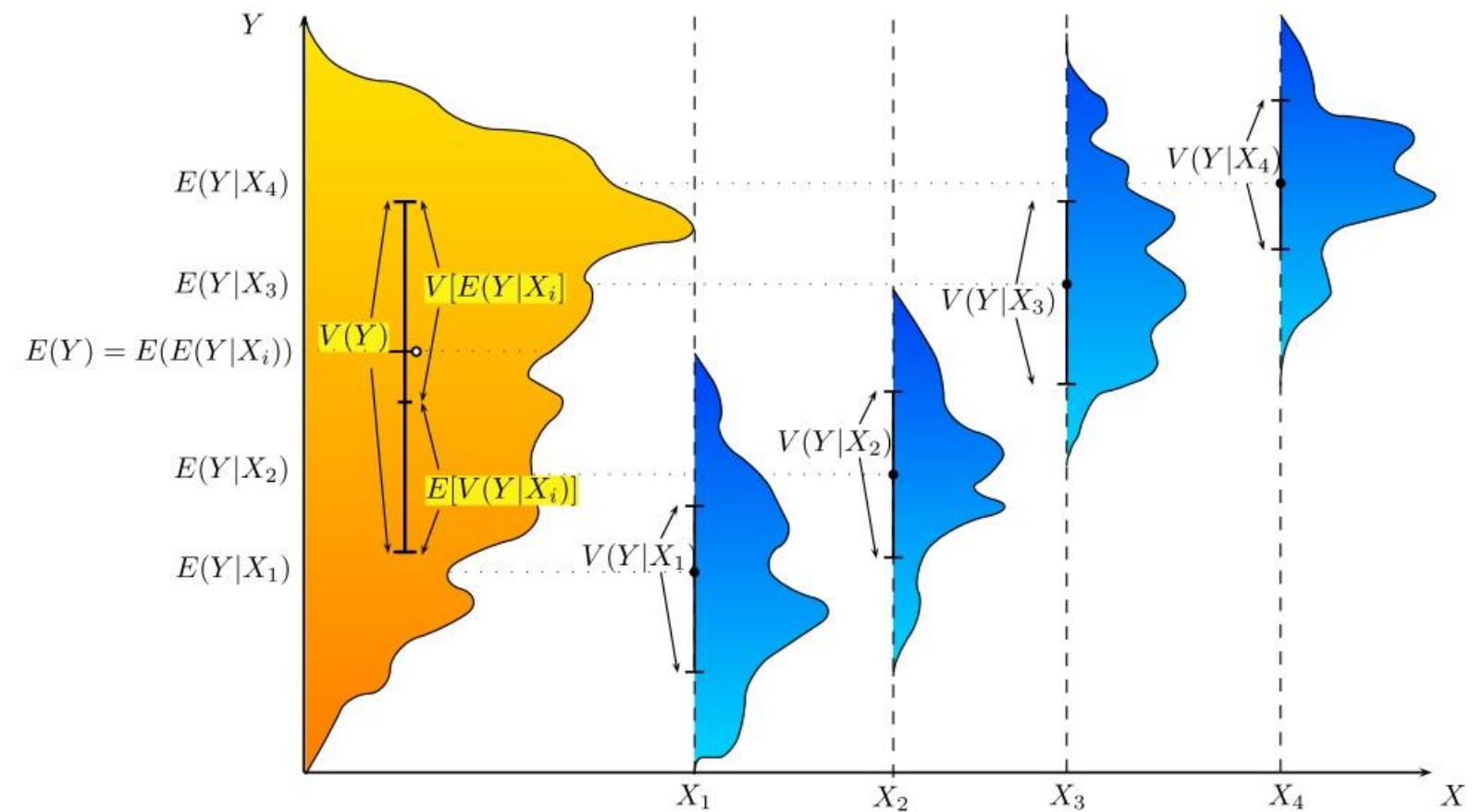
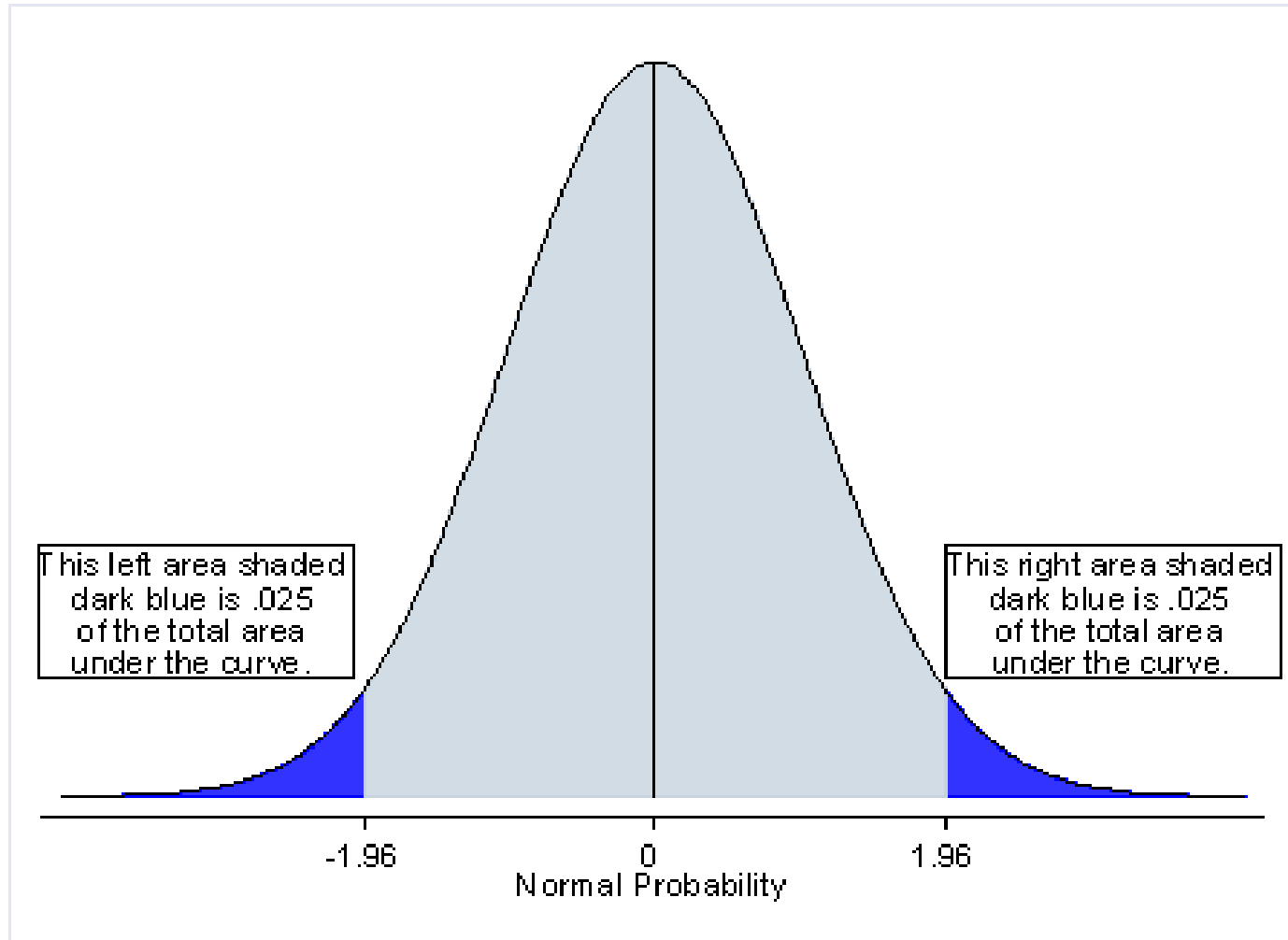
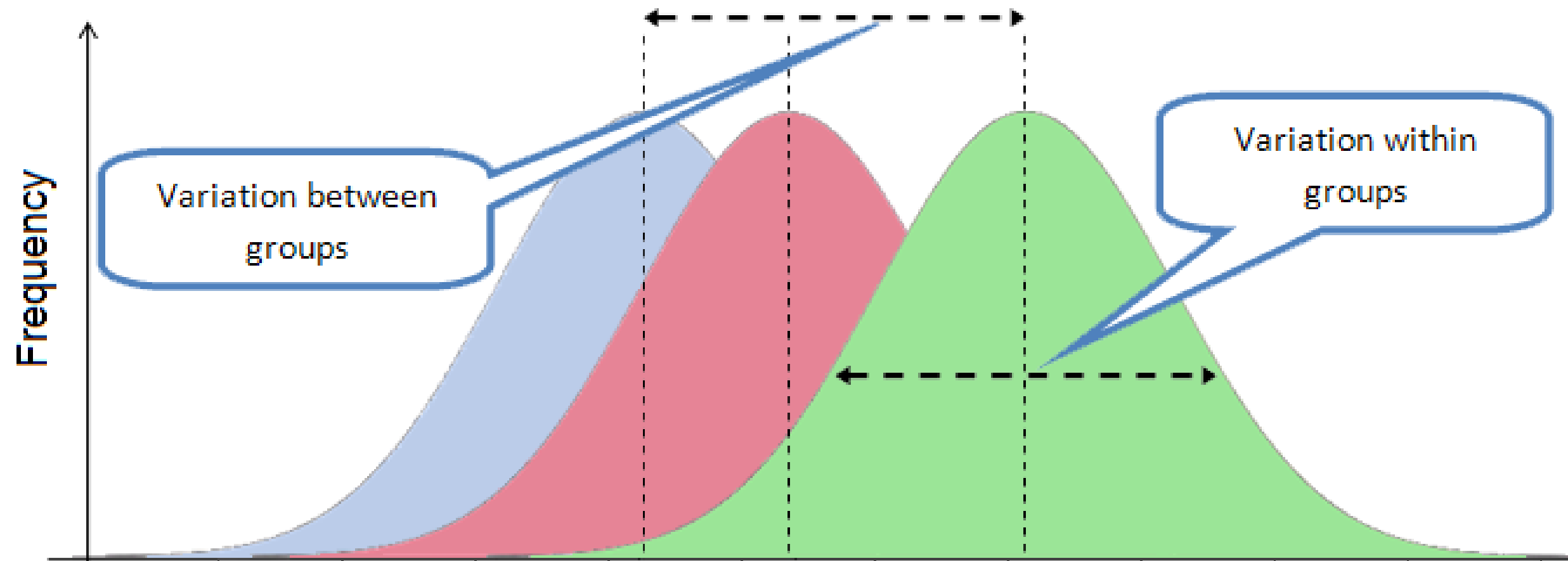


Figure 1: ANOVA : Fair fit

http://en.wikipedia.org/wiki/Analysis_of_variance

ANOVA

Hipótesis Iniciales de Normalidad y Homogeneidad de varianzas.
¿Qué estudia el ANOVA?



ANOVA

```
>summary(fit1)
```

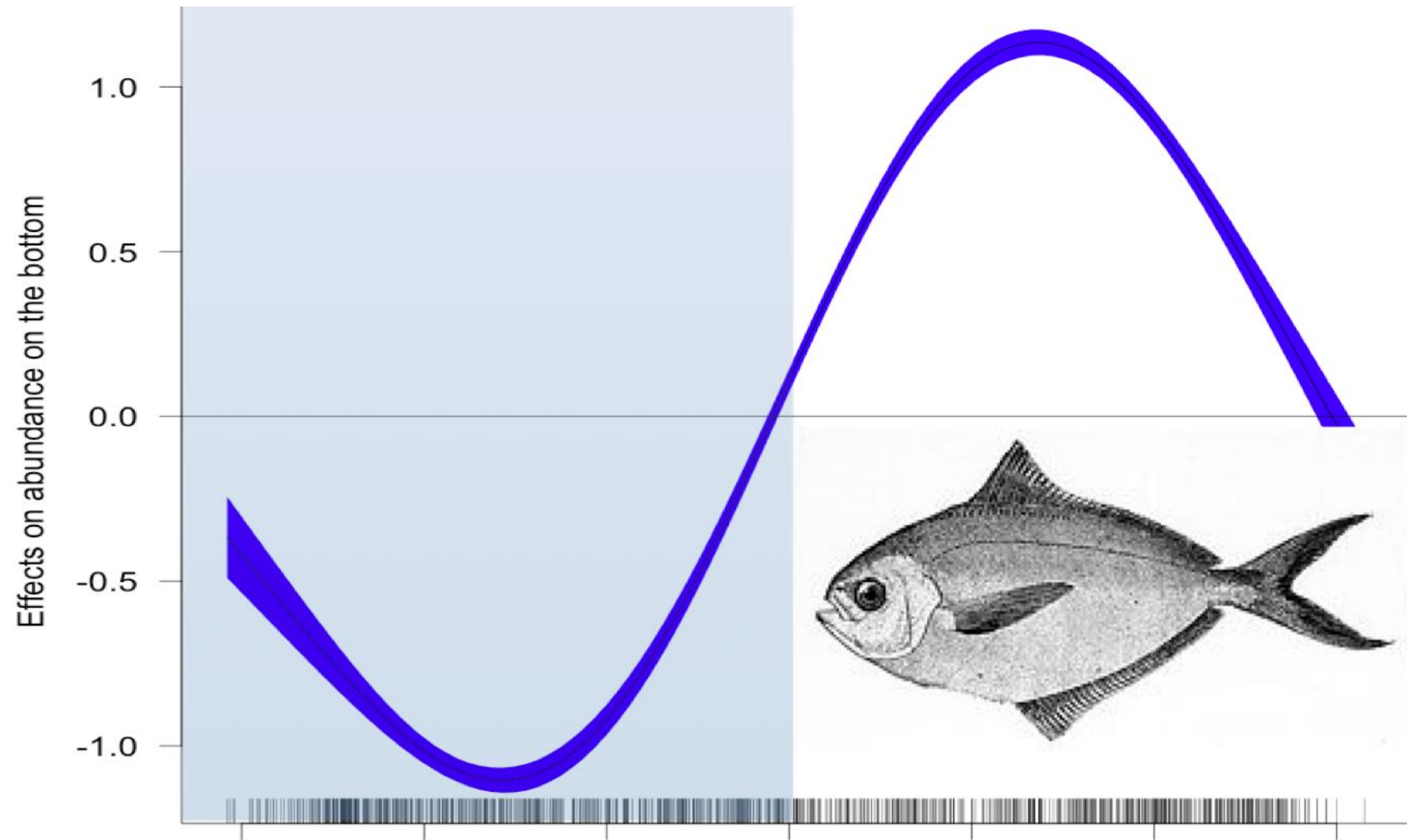
	Df	Sum-Sq	Mean-Sq	F-value	Pr(>F)	
factor(sex)	1	518094	518094	22.622	4.82e-06	***
factor(species)	2	11764	5882	0.257	0.7739	
f(sex):f(species)	2	217443	108722	4.747	0.0101	*
Residuals	141	3229205	22902			

¿Es mi ANOVA válido?

- Homocedasticidad de los Residuos.
- Normalidad de los residuos.
- Los residuos deben de estar Centrados en 0.
- Los residuos no debes estar correlados.

Consecuencias de un mal uso del ANOVA

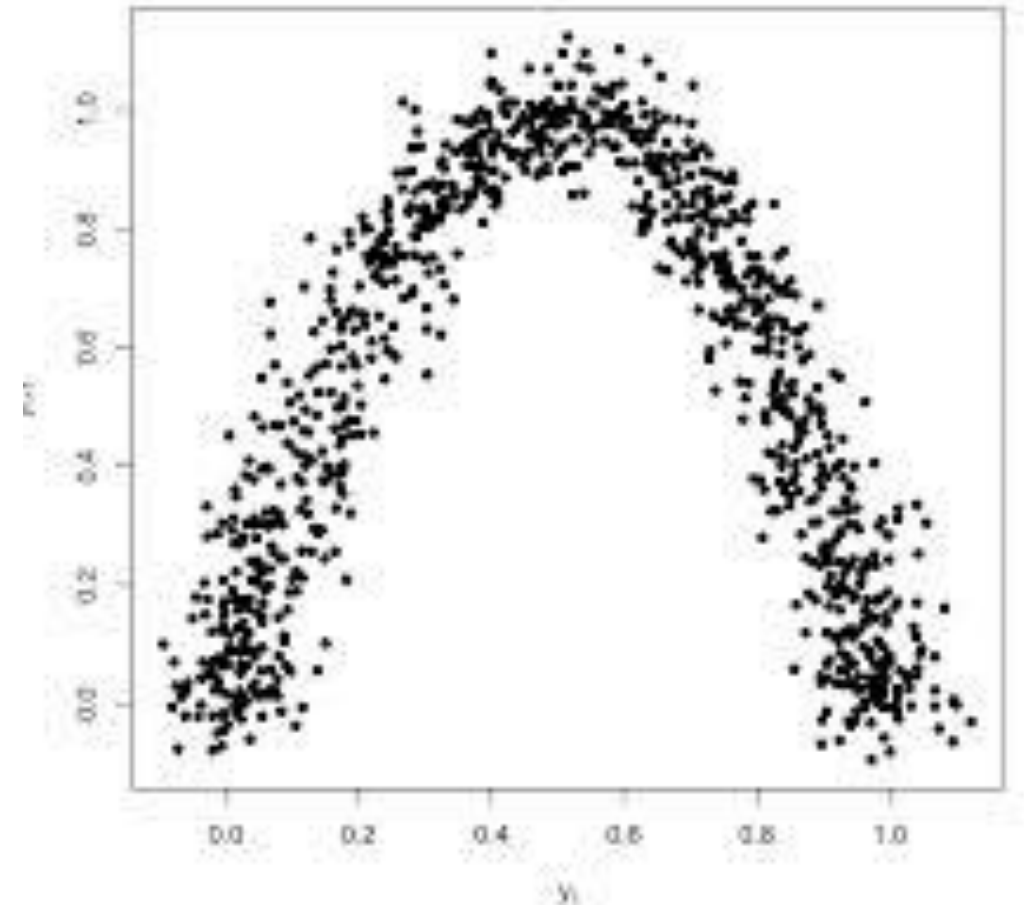
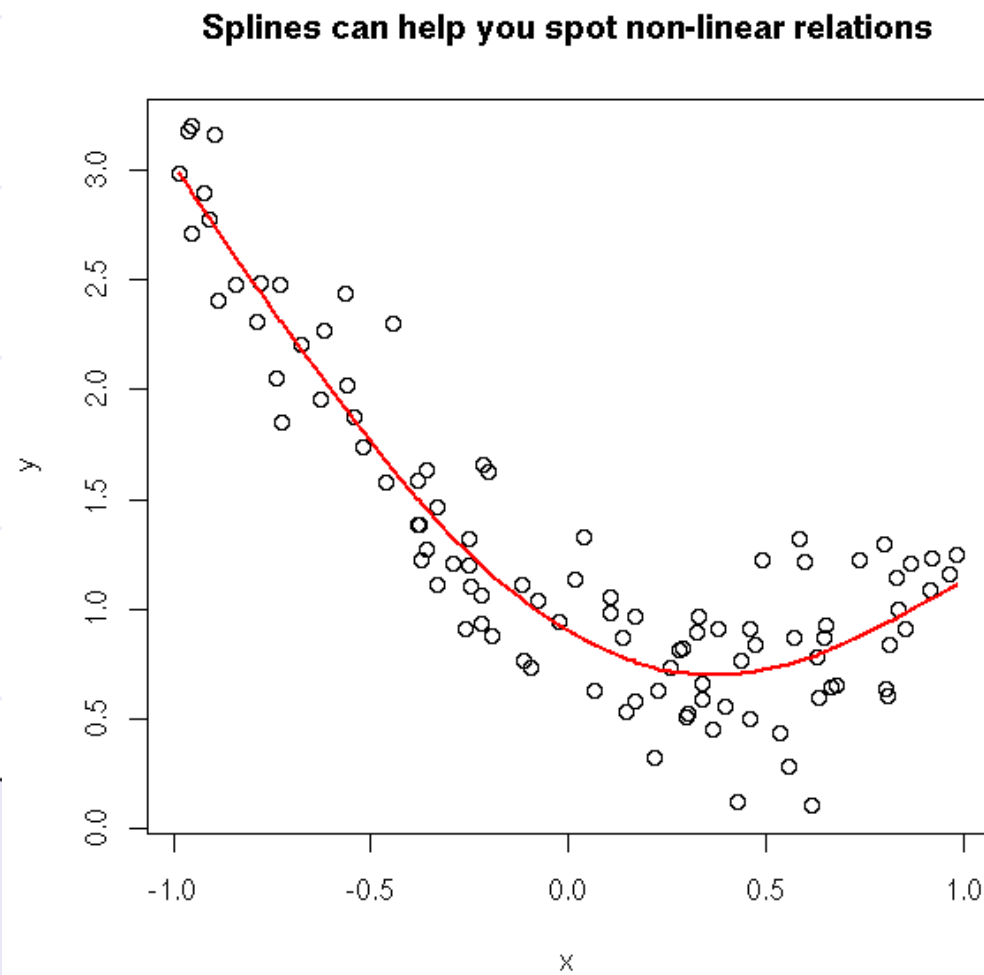
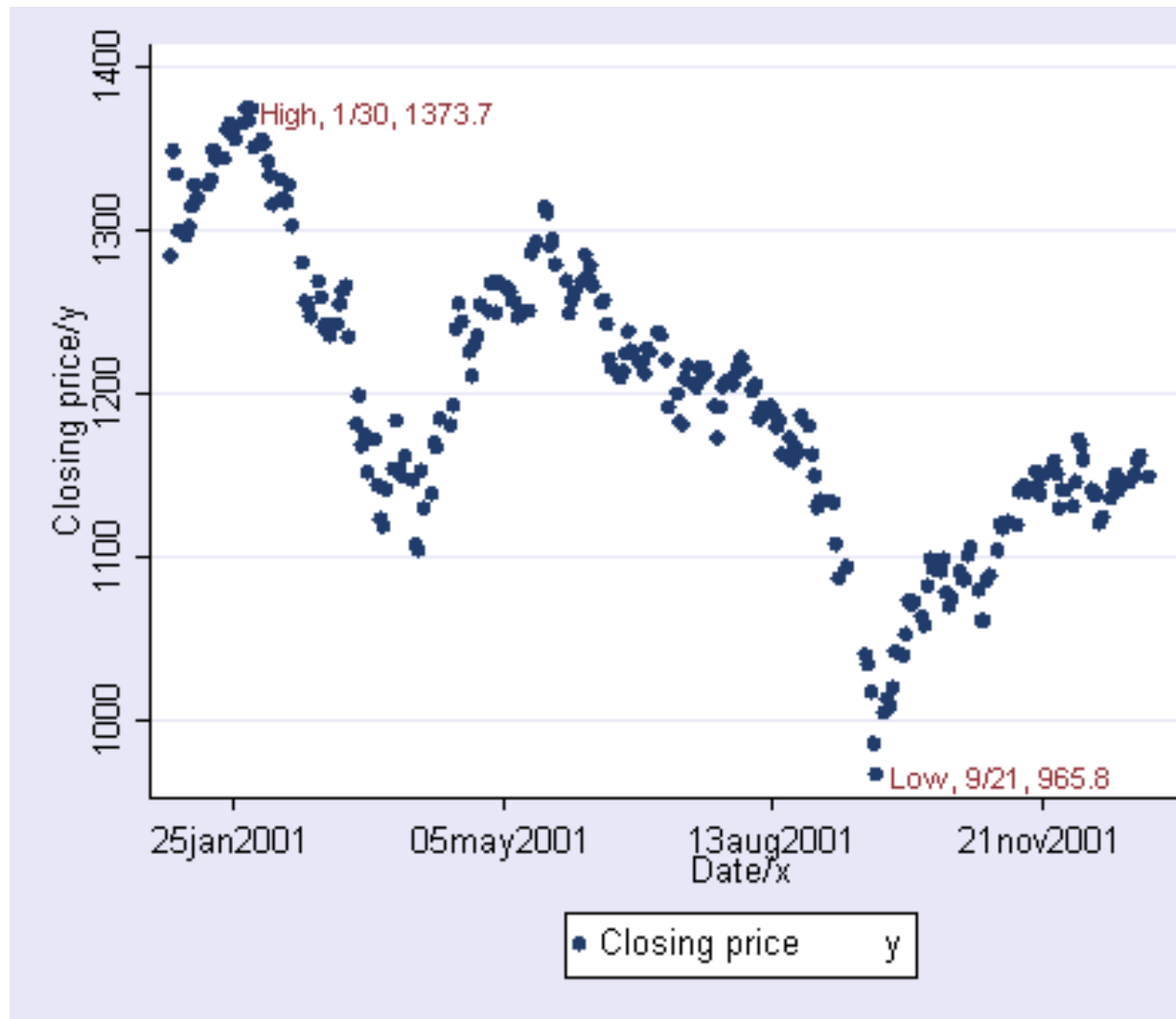
- Realizar más experiencias de las necesitas → \$\$.
- Obtener conclusiones sobre un producto que luego no se cumplen en el mercado → \$\$\$
- **¿Y el prestigio de la Empresa o Centro de Investigación?**



Más allá del ANOVA, la no linealidad.

Modelos Aditivos Generalizados y otros

Séptima Pregunta



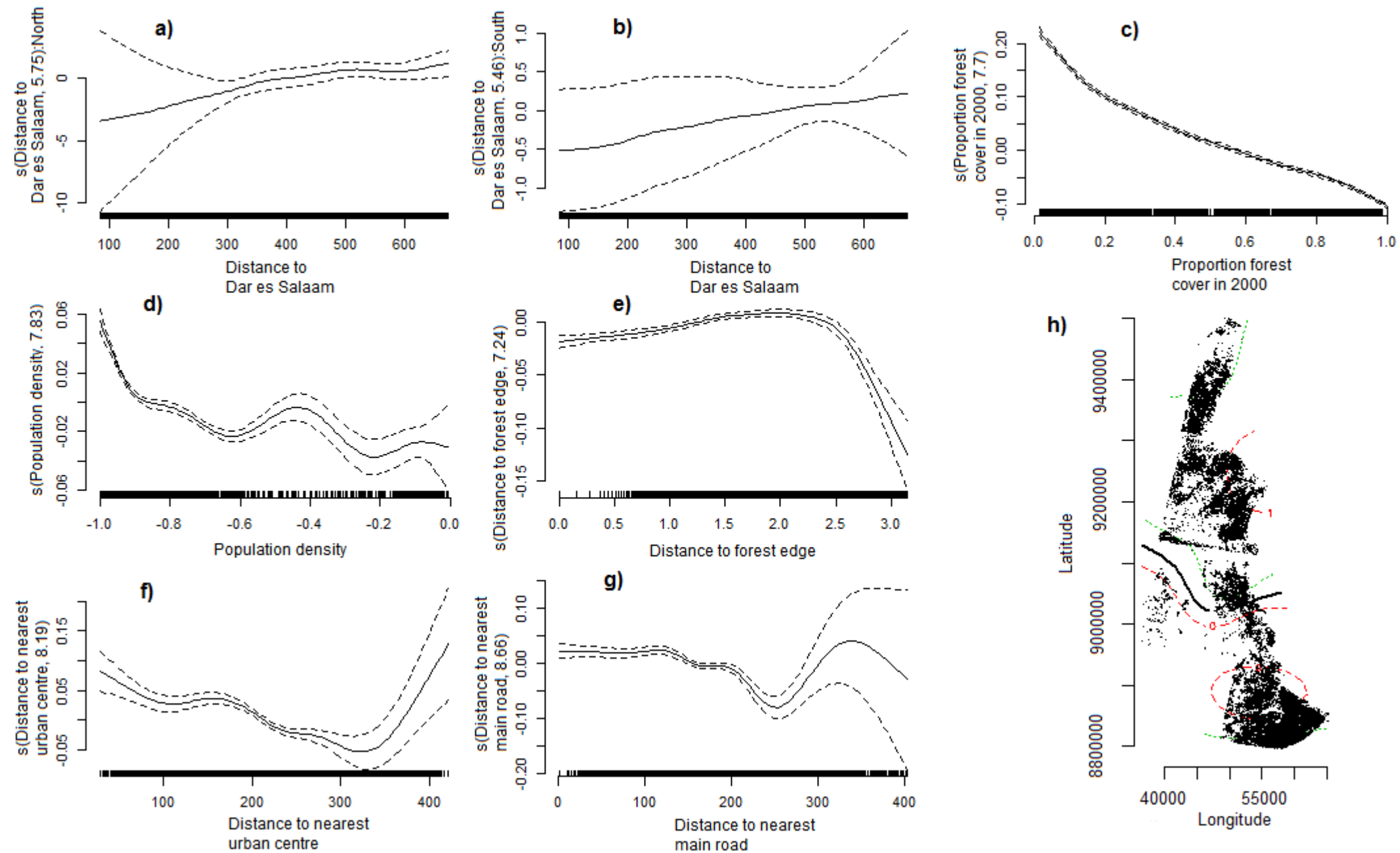
Hay vida más allá del ANOVA

1. Modelos Lineales (Análisis de la Covarianza)
2. Modelos Lineales Mixtos
3. Modelos Lineales Generalizados
4. Partial Least Square
5. Modelos Aditivos Generalizados
6. Etc...

Modelos Aditivos Generalizados (GAM)

- Se trata de realizar como si de un ANOVA o ANCOVA se tratase, pero además la no linealidad ya no nos afecta.
- Ejemplo:
 - Para experimentos donde se producen crecimientos microbianos de forma exponencial,
 - para experimentos donde conforme avanza el tiempo la degeneración de un producto es mayor,
 - etc.

Modelos Aditivos Generalizados

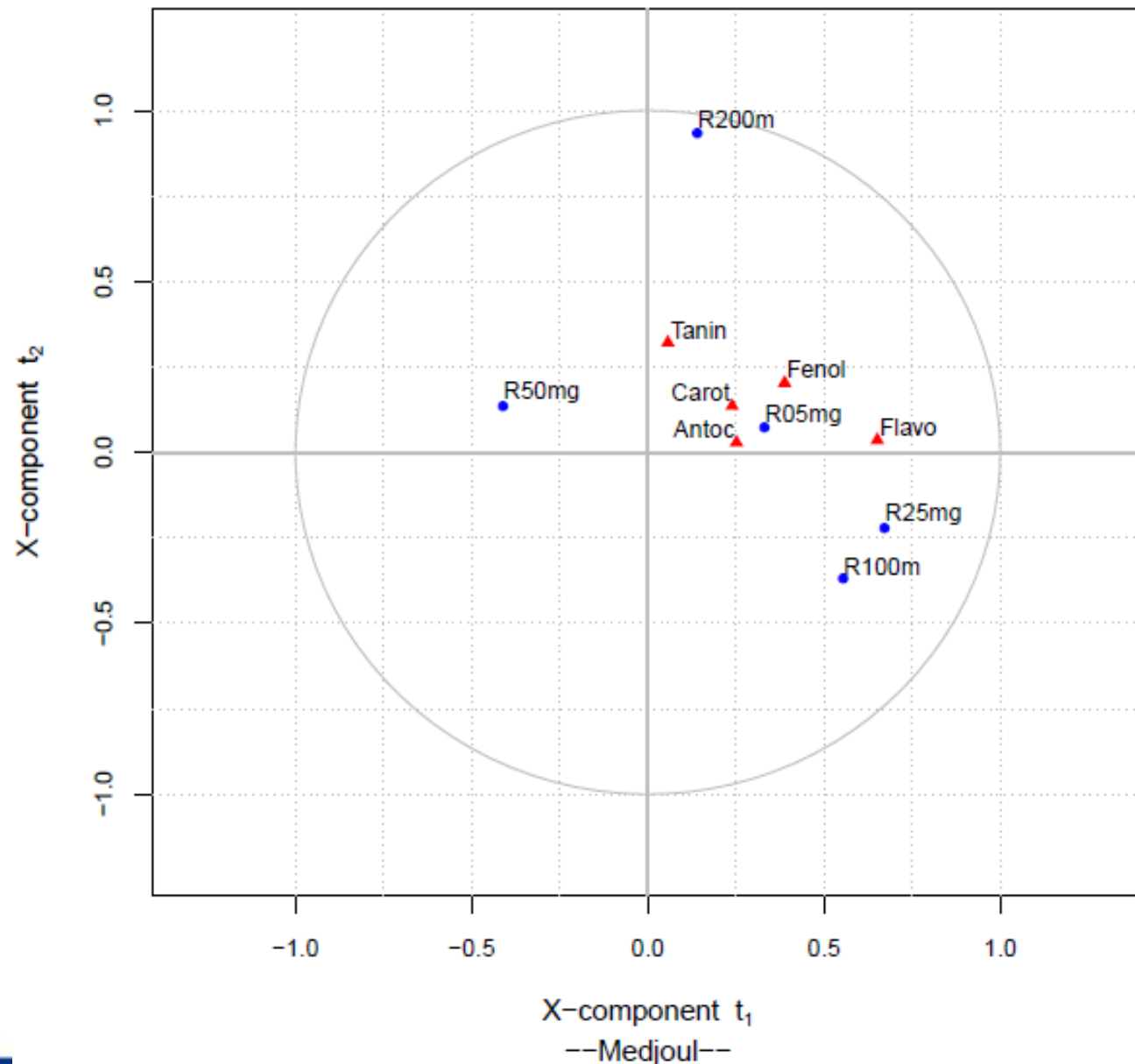


Hay vida más allá del ANOVA (II)

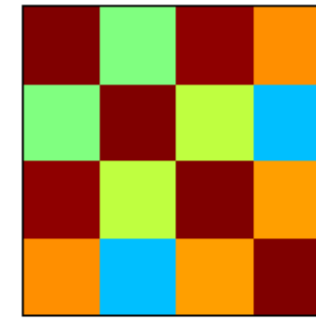
- Métodos Multivariantes:
 - Componentes Principales y Análisis Factorial
 - Análisis Cluster y Multidimensional
 - Árboles de Decisión
 - Redes Neuronales
 - Análisis de Correspondencias
 - Multivariate Partial Least Square

Partial Least Square -2

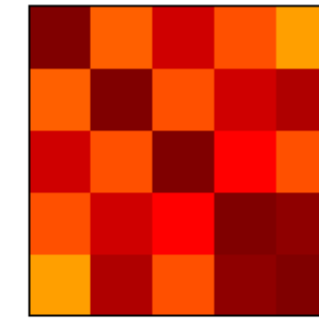
Circle of Correlations on t_1, t_2



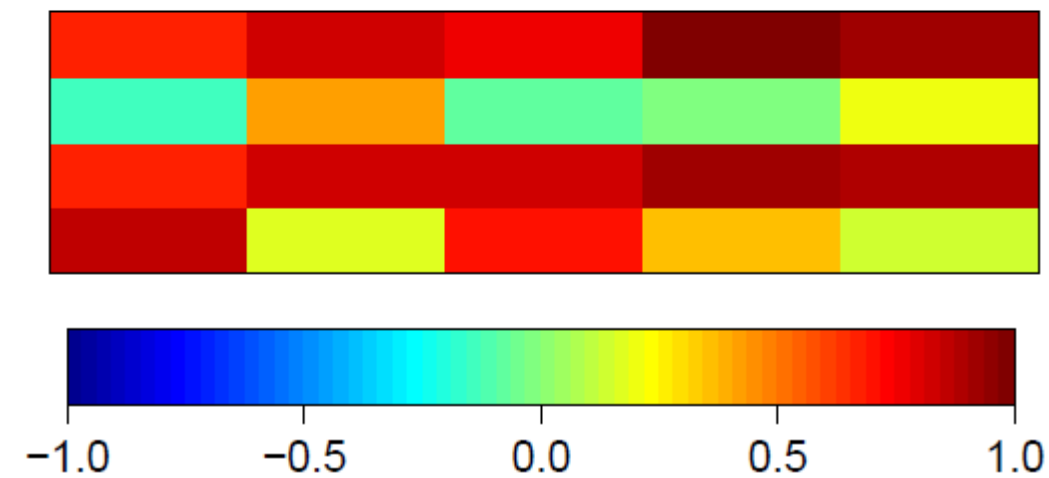
X correlation

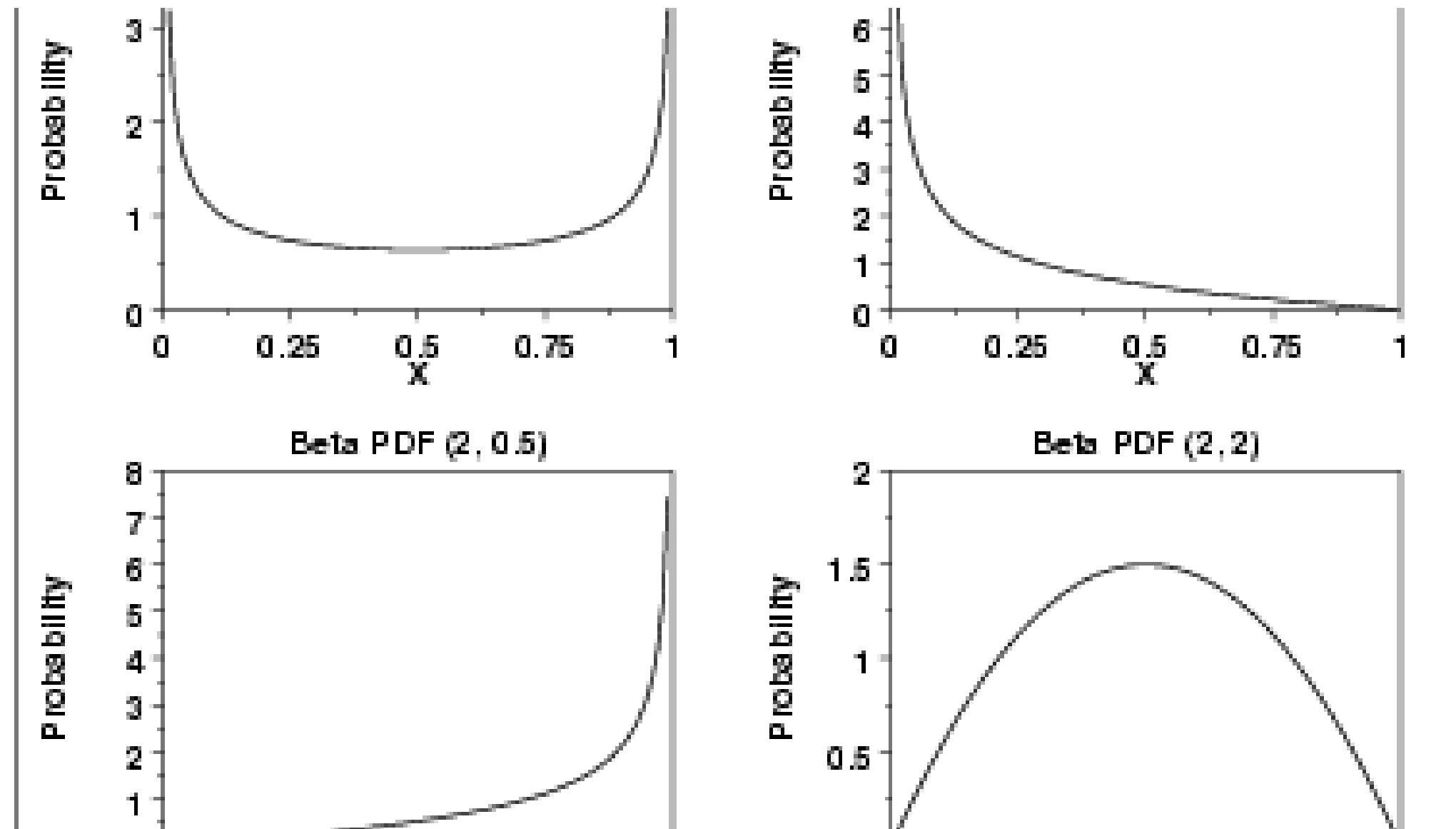


Y correlation



Cross-correlation



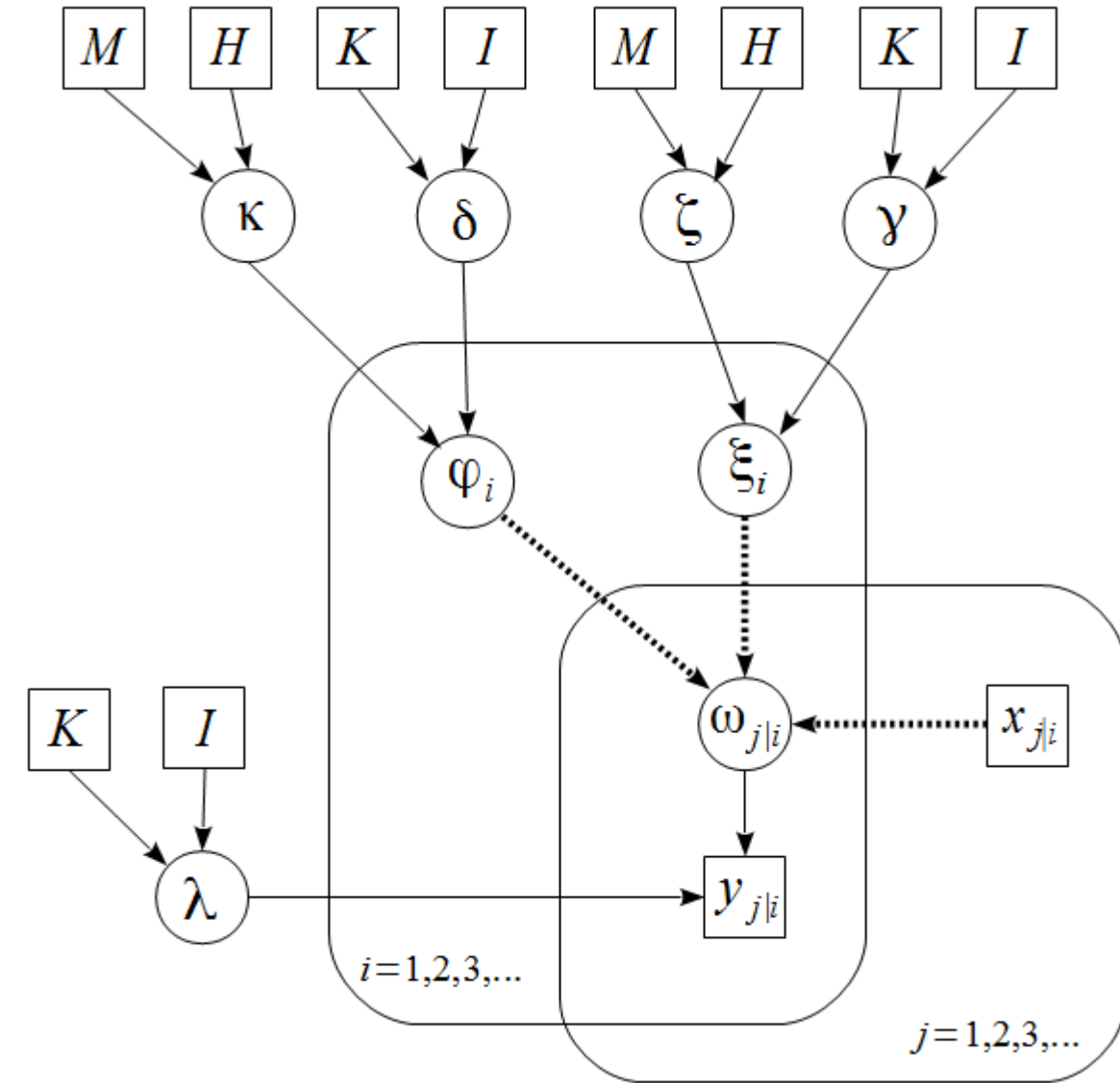
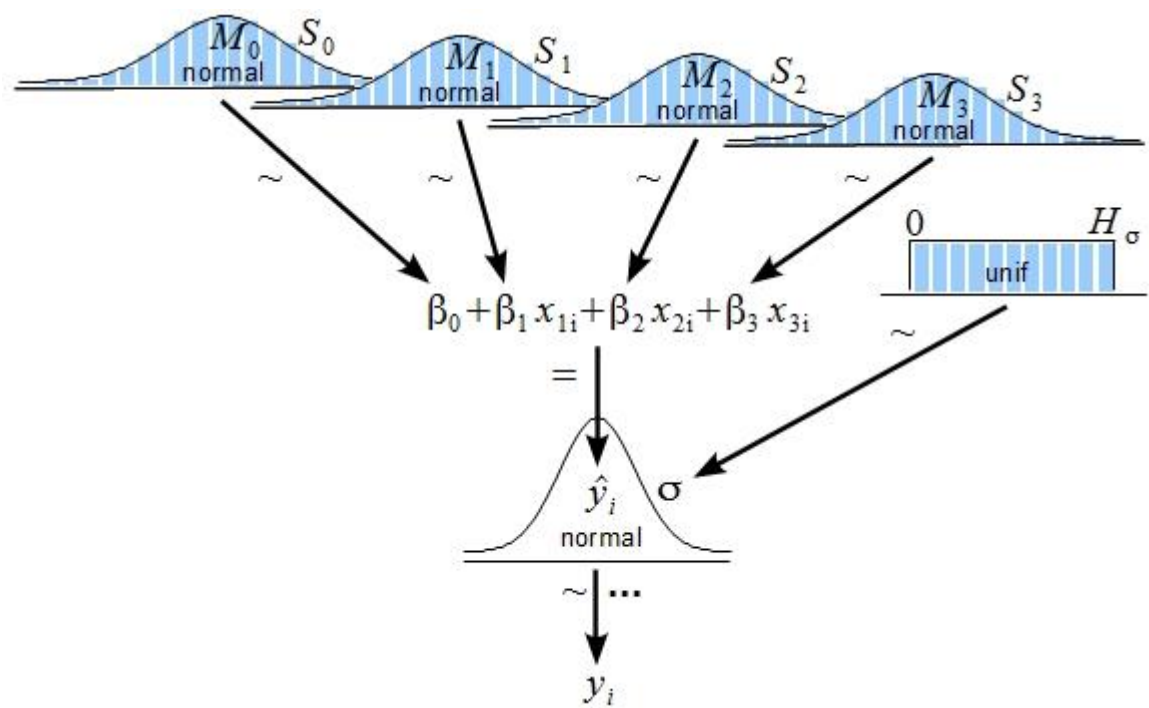


El diseño de experimentos como problema de decisión

Lo que me puede aportar al estadística Bayesiana

Octava Pregunta

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$



Ventajas de la Est. Bayesiana

Necesita menos muestras → **-\$\$\$**

Se puede aprovechar la información disponible que se tenga a priori.

Los modelos son más “naturales”. Podemos expresar las jerarquías y los efectos del tiempo de una forma más plana, los valores de las medias a contrastar, etc.

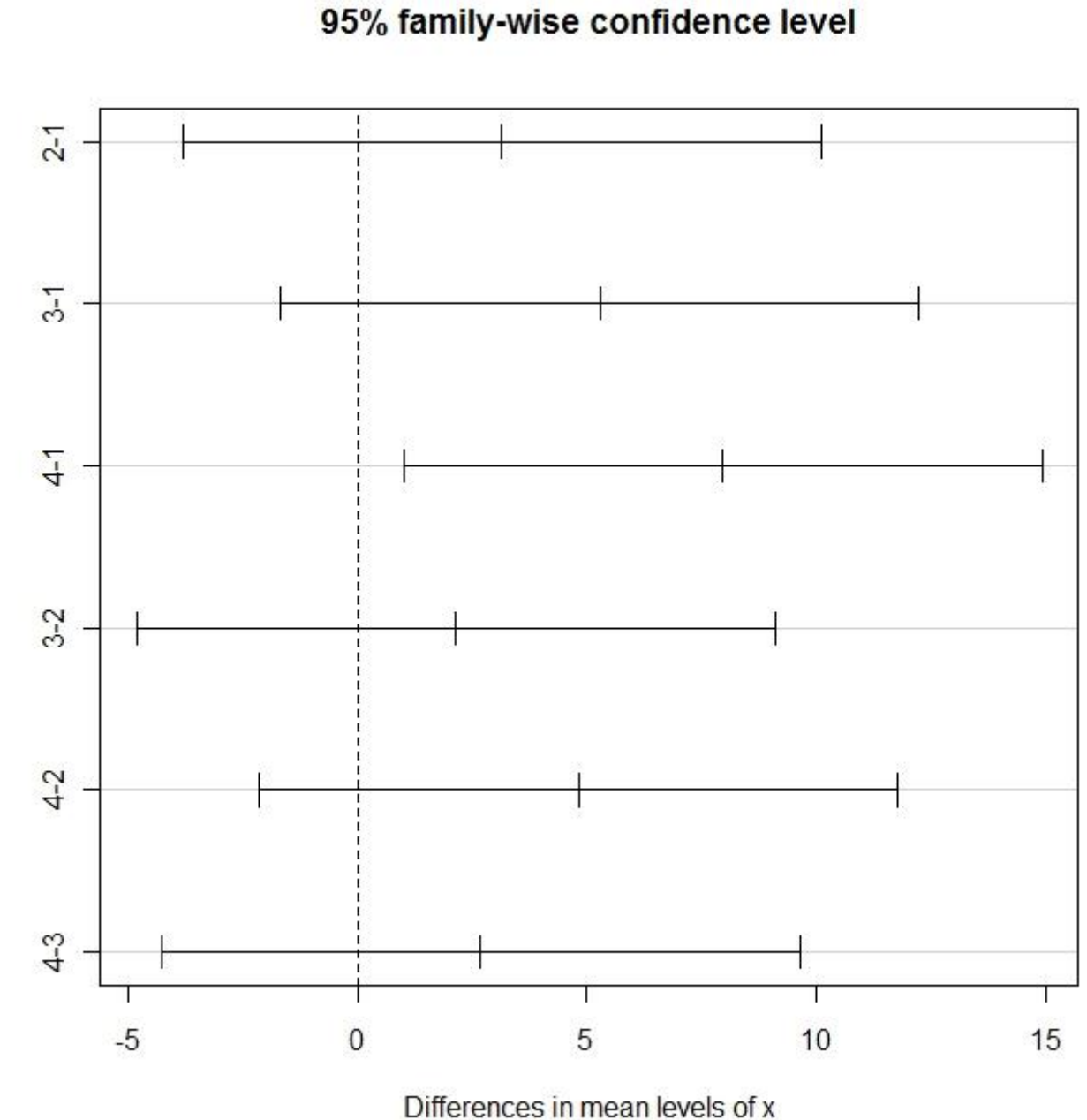
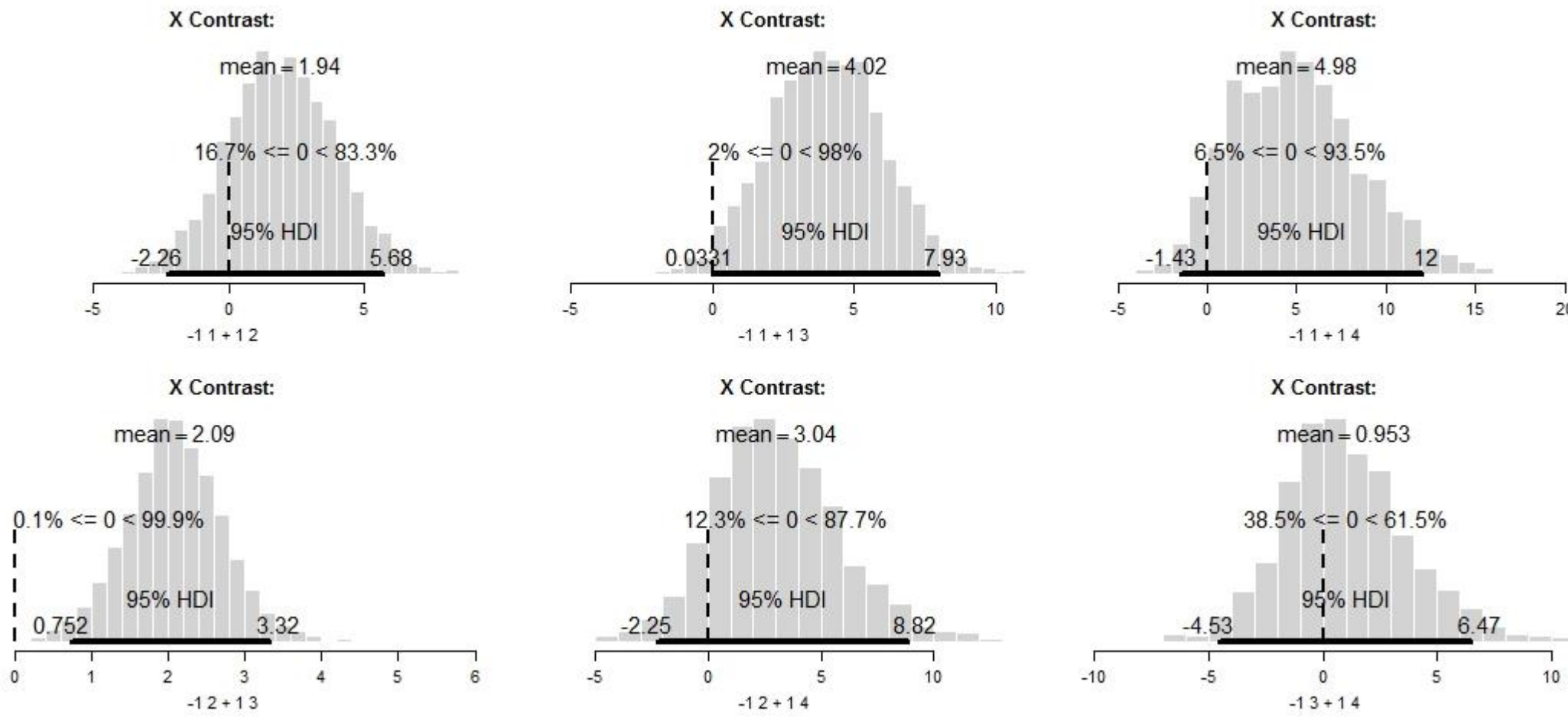
Las conclusiones son en términos de Probabilidades.

¿ANOVA bayesiano?

Sí, en la actualidad existe una réplica de casi todos los análisis “clásicos” en el entorno bayesiano.

Y las conclusiones en un “ANOVA” bayesiano son similares, pero dejando a un lado la estricta norma del $p\text{-value} \leq 0,05$.

ANOVA Bayesiano



Conclusión

- Documentarse bien,
- Tener un tamaño de muestra Adecuado,
- Realizar una buena depuración y análisis descriptivo de los datos,
- Utilizar la Técnica Adecuada para los datos,
- Tener la certeza de la reproducibilidad.

Muchas Gracias

Profesor: Xavier Barber Vallés
I.U.I. Centro de Investigación Operativa
Universidad Miguel Hernández de Elche
Valencia Bayesian Research Group
<http://vabar.github.io/>

