Reducción: diseño experimental y determinación del tamaño de muestra

Tomás Redondo
Comité de Etica de Experimentación Animal - EBD

25/04/2017

Reducción:

“Ways of obtaining comparable levels of information from the use of fewer experimental animals, or of obtaining more information from a given number of animals, so that fewer animals are needed to complete a given research project, taking into account individual animal welfare in relation to minimising pain, suffering, distress or lasting harm”

ECVAM - European Centre for the Validation of Alternative Methods

Existe una clara relación entre Reducción y Calidad científica

Un proyecto de mala calidad:

- **No realiza una contribución significativa al conocimiento.**
- **No consigue lograr sus objetivos.**
- **Es incierto si ha alcanzado los objetivos o no.**
- **No puede ser replicado.**
- **Se hubiera alcanzado el mismo logro con un menor número de animales.**

Existe una clara relación entre Reducción y Calidad científica

Un proyecto de mala calidad:

Desperdicia tiempo
(Algo que no nos sobra a los científicos y que perjudica la reputación)

Desperdicia dinero y recursos
(Algo que debería enfadar a quien da el dinero y al instituto)

Viola los principios éticos si utiliza animales o humanos o daña al medio ambiente:

- **Debe usar el número mínimo de animales**
- **Debe tener una expectativa de éxito razonable**

5. **Aplicación de métodos para reemplazar, reducir y refinar el uso de animales en procedimientos. Justificación:**

Métodos de reemplazamiento	**P1/P2: no** **P3: Uso de animales domésticos**
Métodos de reducción	**P1: N1 (……)** **P2: N2 (……)** **P3: (……)**
Métodos de refinamiento	**IDENTIFICAR ALTERACIONES DEL BIENESTAR PROVOCADAS POR LAS INTERVENCIONES Y LOS METODOS QUE SE EMPLEARAN PARA MINIMIZARLAS**

Existe una clara relación entre Reducción y Calidad científica

Un proyecto de mala calidad:

Desperdicia tiempo
(Algo que no nos sobra a los científicos y que perjudica la reputación)

Desperdicia dinero y recursos
(Algo que debería enfadar a quien da el dinero y al instituto)

Viola los principios éticos si utiliza animales o humanos o daña al medio ambiente:

- **Debe usar el número mínimo de animales**
- **Debe tener una expectativa de éxito razonable**

Es ilegal

Experimento:

“A planned strategy for collecting a set of data to answer a specific question”

Festing (2010)
UFAW Handbook on the Care and Management of Laboratory and Other Research Animals

Tipos de experimentos

Estudio piloto

- Puesta a punto de la logística, información preliminar

Experimento exploratorio

- El objetivo es generar hipótesis

- Puede no existir una predicción clara

- Se miden muchas variables

- El análisis estadístico es problemático

Experimento confirmatorio

- **Hipótesis formal a priori**
- **Análisis estadístico claro. Conduce a decisiones en base a pruebas estadísticas, por lo que los criterios (por ejemplo, valores de *P*) deben ser fiables.**

Tipos de experimentos

Estudio piloto

- Puesta a punto de la logística, información preliminar

Experimento exploratorio

- El objetivo es generar hipótesis

- Puede no existir una predicción clara

- Se miden muchas variables

- El análisis estadístico es problemático

Experimento confirmatorio

- **Hipótesis formal a priori**
- **Análisis estadístico claro. Conduce a decisiones en base a pruebas estadísticas, por lo que los criterios (por ejemplo, valores de *P*) deben ser fiables.**

Diseño experimental:

Es la parte de la estadística que usamos ANTES de hacer el experimento para optimizar la información en función de los recursos

Diseño experimental:

Es la parte de la estadística que usamos ANTES de hacer el experimento para optimizar la información en función de los recursos

- **El diseño experimental no es malgastar el tiempo. Al contrario, ahorra mucho tiempo despues**
- **Hay errores de muestreo que la estadística no puede corregir a posteriori (p. ej. falta de independencia, controles inapropiados)**
- **No es cierto que si hay una gran cantidad de datos:**
    - **Algo interesante acabará por salir**
    - **Será posible detectar incluso efectos muy sutiles**

Tres pasos importantes que nadie debería saltarse:

1. **<u>Escribir</u> los objetivos concretos del experimento: qué quiero saber.**
2. **Diseñar una estrategia para alcanzar el objetivo:**
    - **Cuantos tratamientos**
    - **Cuantas réplicas**
    - **Cómo se analizarán los datos después**
3. **<u>Apuntar</u> los detalles operativos:**
    - **Cómo se aplicarán los tratamientos**
    - **Cómo se aleatorizará el muestreo**
    - **Cómo se estructurará (en bloques, factorial, etc)**
    - **Cuánto tiempo llevará**
    - **¿Se puede parar a la hora de comer?**
    - **Etc.**

OBJETIVOS DEL DISEÑO EXPERIMENTAL

1. Evitar artefactos experimentales
2. Eliminar sesgos
    - Definiendo grupo control
    - Aleatorización
    - Diseño ciego
3. Reducir el error de muestreo
    - Independencia estadística de las réplicas
    - Bloques y factores
    - Balance

1. **Evitar artefactos experimentales**

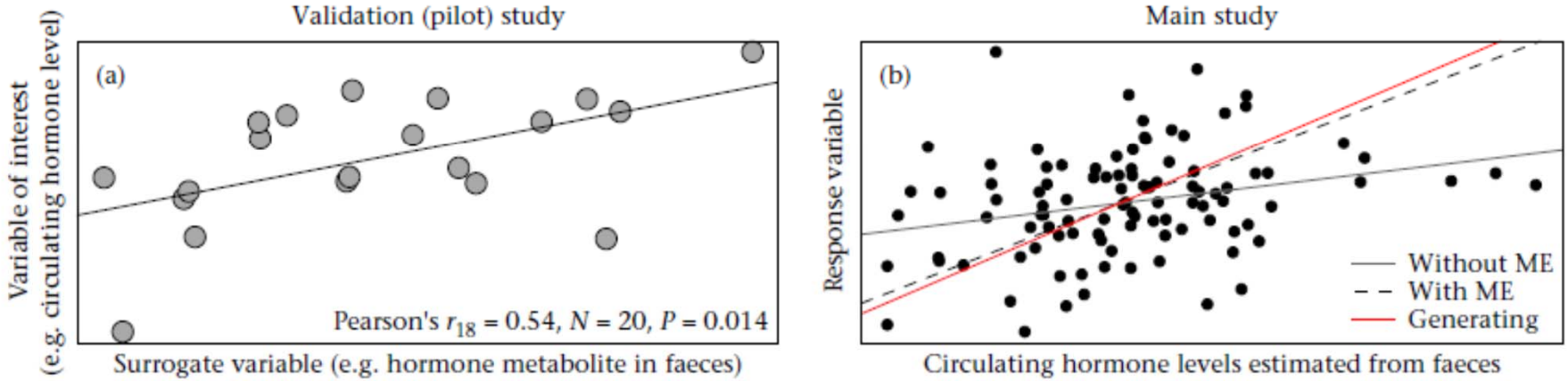
“Un sesgo en una medida producido por efectos no deseados del procedimiento experimental”

- **Condiciones artificiales**
- **Material de medida inadecuado**
- **Personal no entrenado**
- **Repetibilidad**
- **Fiabilidad de variables indirectas y *proxies***

1. Evitar artefactos experimentales

“Un sesgo en una medida producido por efectos no deseados del procedimiento experimental”

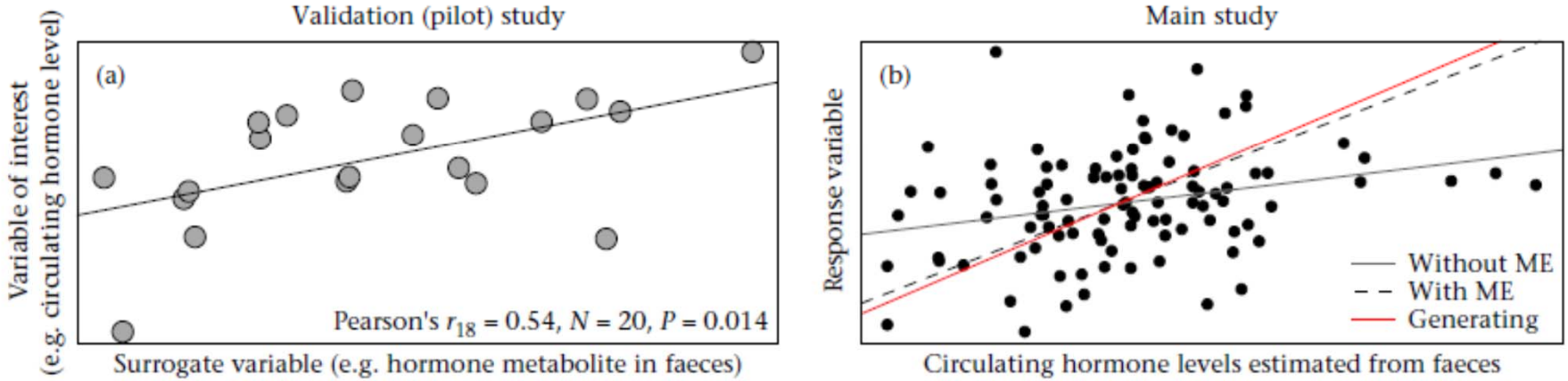
- **Condiciones artificiales**
- **Material de medida inadecuado**
- **Personal no entrenado**
- **Repetibilidad**
- **Fiabilidad de variables indirectas y *proxies***



1. Evitar artefactos experimentales

“Un sesgo en una medida producido por efectos no deseados del procedimiento experimental”

- **Condiciones artificiales PILOTO**
- **Material de medida inadecuado PILOTO**
- **Personal no entrenado PILOTO**
- **Repetibilidad PILOTO**
- **Fiabilidad de variables indirectas y *proxies* PILOTO**

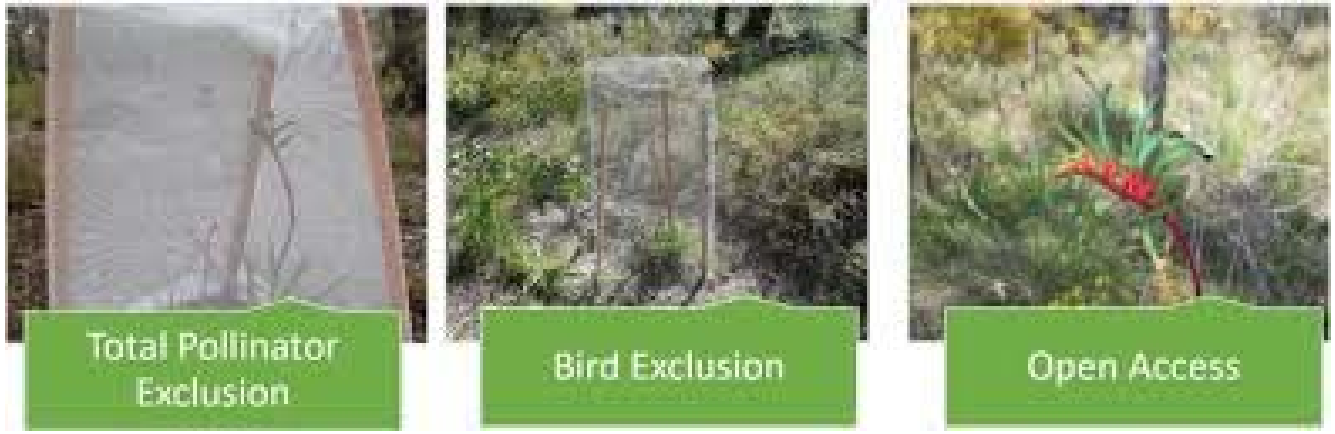


1. Evitar artefactos experimentales
2. **Eliminar sesgos**

- Definiendo grupo control

En iguales condiciones salvo tratamiento. Efecto placebo hasta del 40 %

No siempre es fácil



1. Evitar artefactos experimentales
2. **Eliminar sesgos**
   - Definiendo grupo control
   - **Aleatorización**

Permite utilizar estadísticos muestrales ($\bar{x}$, s^2) para estimar de forma fiable (no sesgada) parámetros (μ, σ^2) de nuestra "*Población*" y sus errores de muestreo

Cualquier unidad de muestreo debe tener una probabilidad igual de ser elegida (independencia)

- Cada nivel de la muestra (p. ej. Experimental/control) representa a la población

Idealmente: https://www.randomizer.org/

En la práctica suele ser caótico (haphazard), no aleatorio (random)

- Cada unidad de muestreo debe tener la misma probabilidad de ser asignada a un tratamiento

1. Evitar artefactos experimentales
2. **Eliminar sesgos**
   - Definiendo grupo control
   - **Aleatorización**

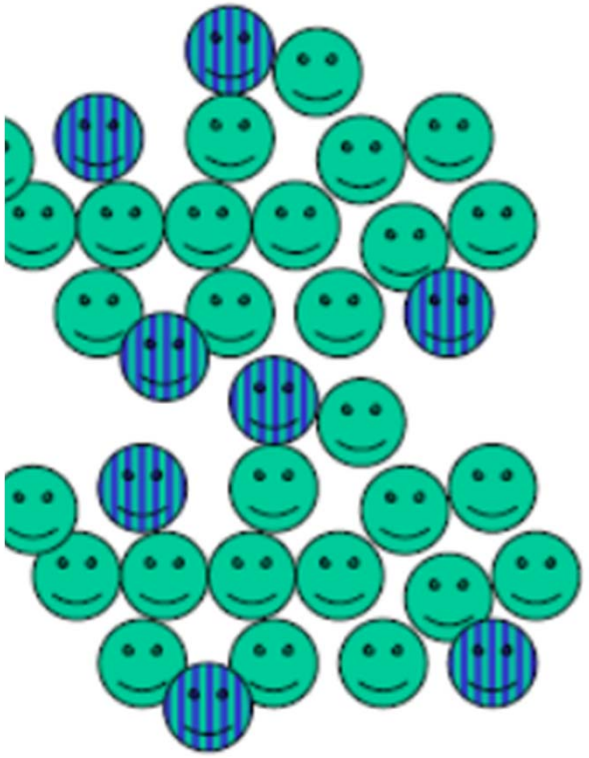
Tipos de muestreo aleatorio

- **Simple**
- **Estratificado**
- **Sistemático**
- **Cluster**

1. Evitar artefactos experimentales
2. **Eliminar sesgos**
   - Definiendo grupo control
   - **Aleatorización**

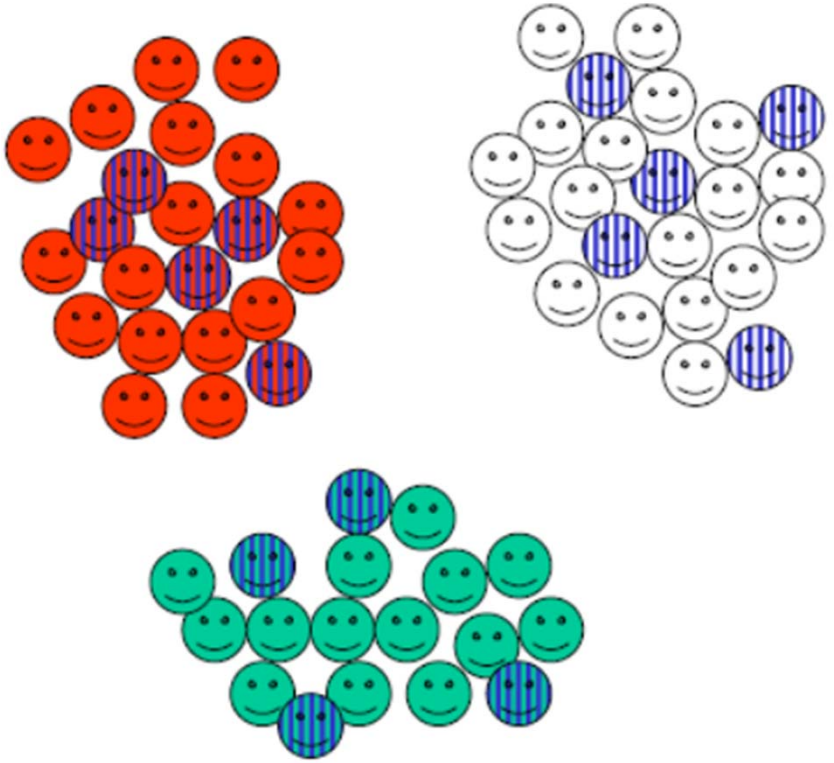
Tipos de muestreo aleatorio

- **Simple**
- Estratificado
- Sistemático
- Cluster



- No precisa de conocimiento previo, fácil de analizar, buena estima de errores
- Mayor error y por tanto mayor tamaño muestral

1. Evitar artefactos experimentales
2. **Eliminar sesgos**
   - Definiendo grupo control
   - **Aleatorización**



Tipos de muestreo aleatorio

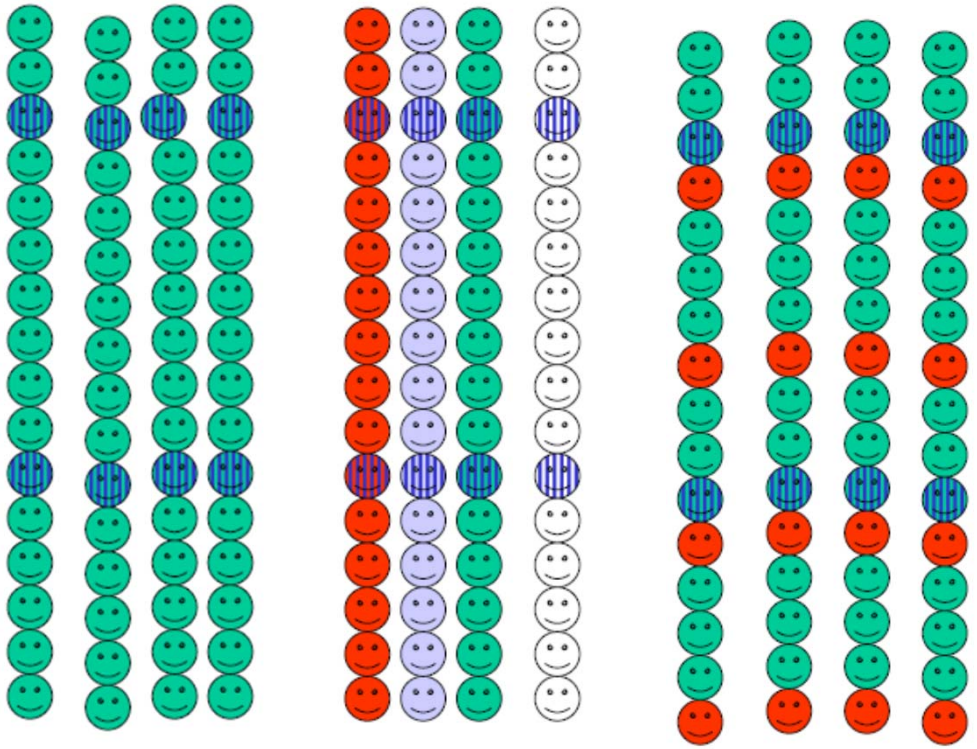
- Simple
- **Estratificado**
- Sistemático
- Cluster

- Representa la estructura de la población, estimas de error más bajas (menor N)
- Requiere información sobre las proporciones de cada estrato

1. Evitar artefactos experimentales
2. **Eliminar sesgos**
   - Definiendo grupo co
   - **Aleatorización**

Tipos de muestreo aleatorio

- Simple
- Estratificado
- **Sistemático**
- Cluster

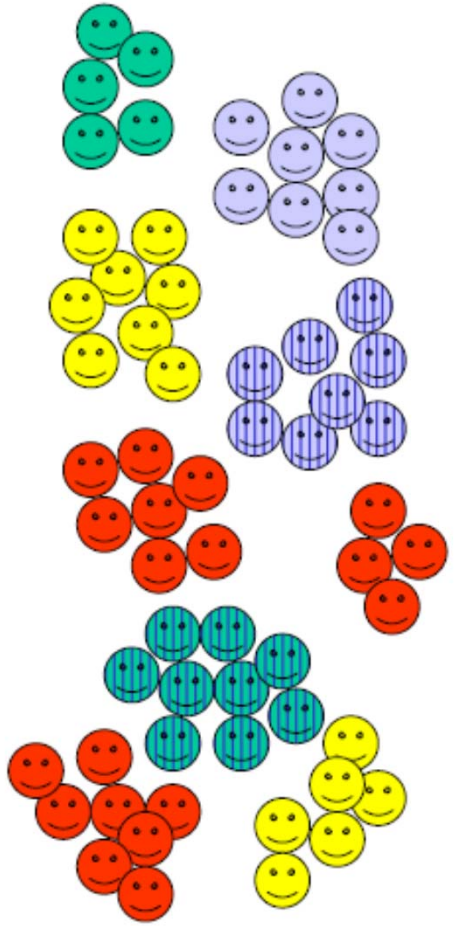


- **Intervalo de muestreo k = N/n. Tras un sorteo que determina j, la serie es j, j+k, j+2k,…**
- **Buena estima de error si la lista es aleatoria, caótica o estratificada**
- **Sujeto a errores si existe estructura periódica**

1. Evitar artefactos experimentales
2. **Eliminar sesgos**
   - Definiendo grupo control
   - **Aleatorización**

Tipos de muestreo aleatorio

- Simple
- Estratificado
- Sistemático
- **Cluster**

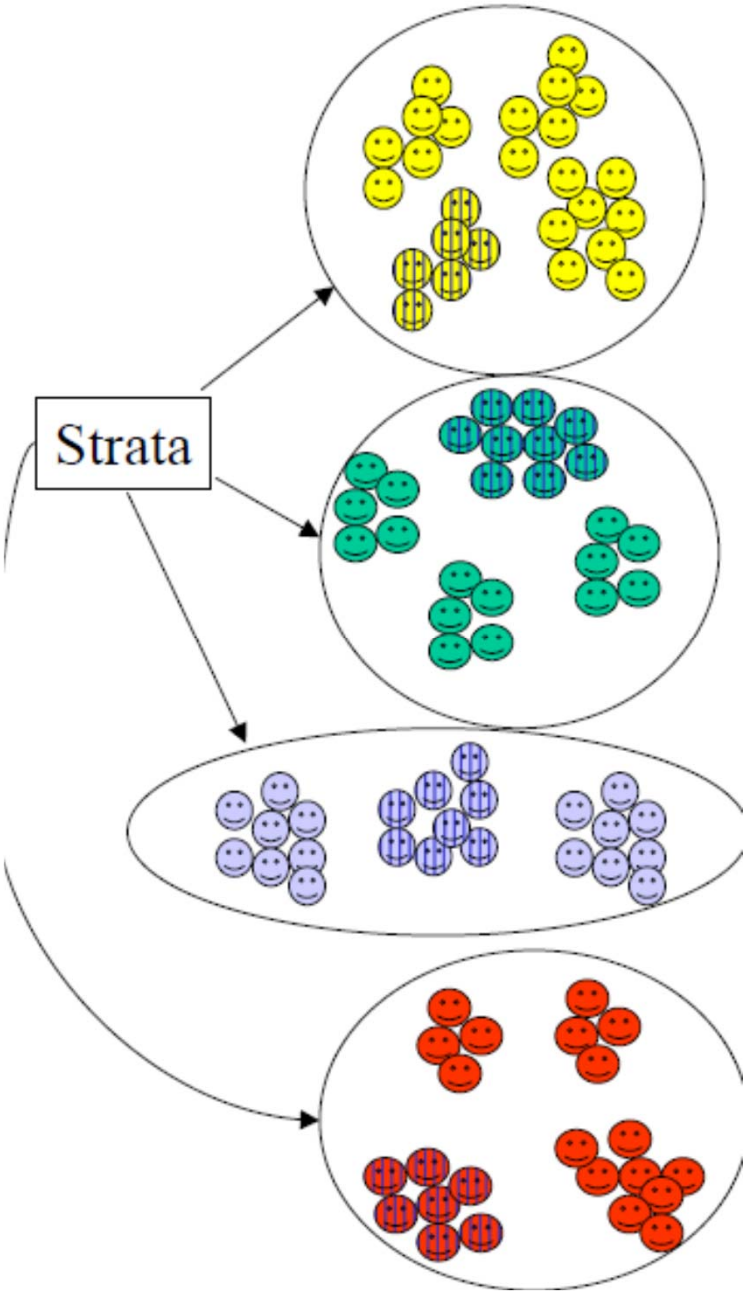


- **Facilidad de aplicación, útil para estimar variación en niveles jerárquicos**
- **(estructura anidada)**
- **Elevado error, no aplicable si los clusters son distintos**

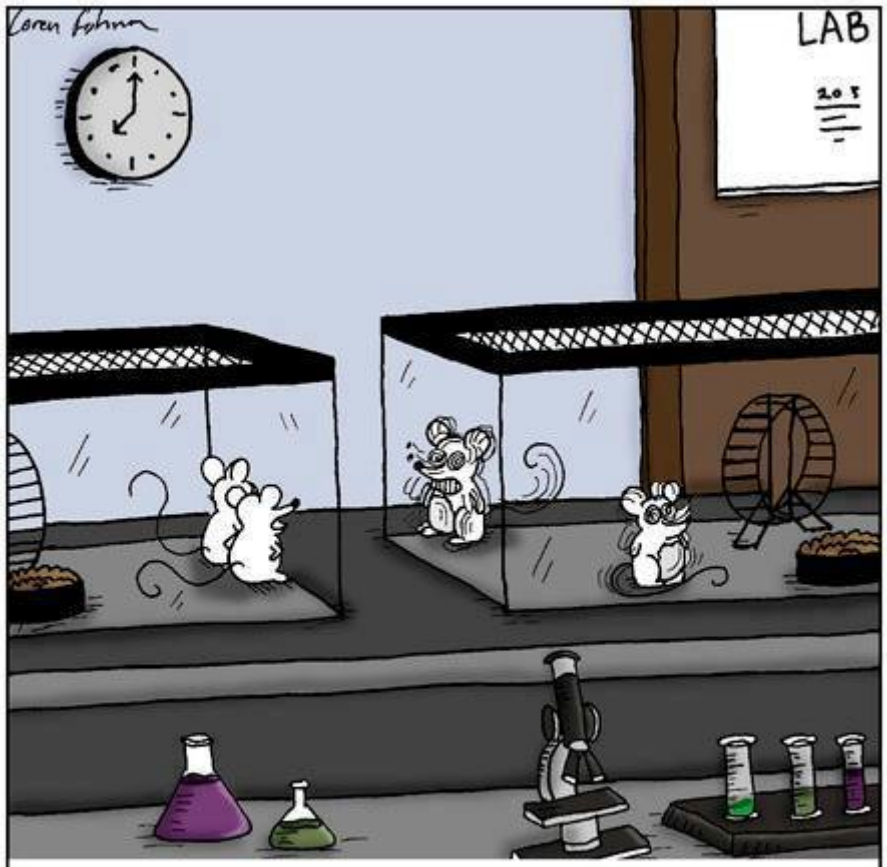
1. Evitar artefactos experimentales
2. **Eliminar sesgos**
   - Definiendo grupo control
   - **Aleatorización**

Tipos de muestreo aleatorio

- Simple
- **Estratificado**
- Sistemático
- **Cluster**



1. Evitar artefactos experimentales
2. Eliminar sesgos
   - Definiendo grupo control
   - Aleatorización
   - **Diseño ciego**



Blind/not blind
odds ratio 3.4 (95% CI 1.7-6.9)

Random/not random
odds ratio 3.2 (95% CI 1.3-7.7)

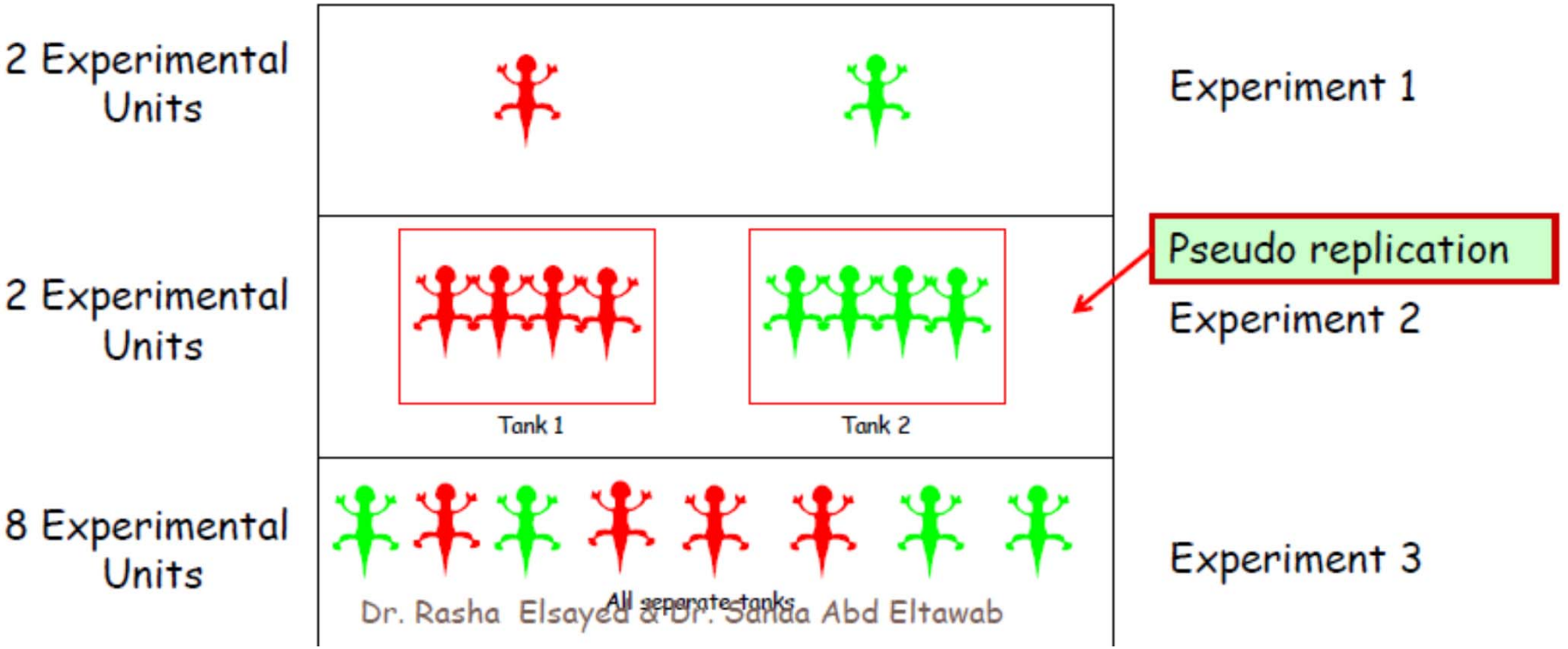
Blind Random/not blind not random
odds ratio 5.2 (95% CI 2.0-13.5)

290 animal studies

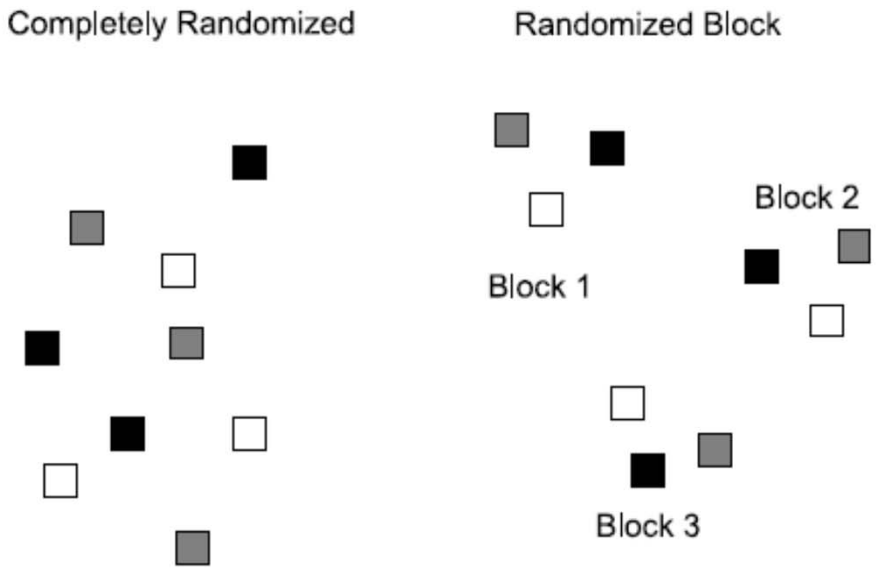
Babasta et al 2003 Acad. emerg. med. 10:684-687

1. Evitar artefactos experimentales
2. Eliminar sesgos
   - Definiendo grupo control
   - Aleatorización
   - Diseño ciego
3. **Reducir el error de muestreo**
   - **Independencia estadística de las réplicas**

$$SE_{\bar{Y}} = \frac{s}{\sqrt{n}}$$



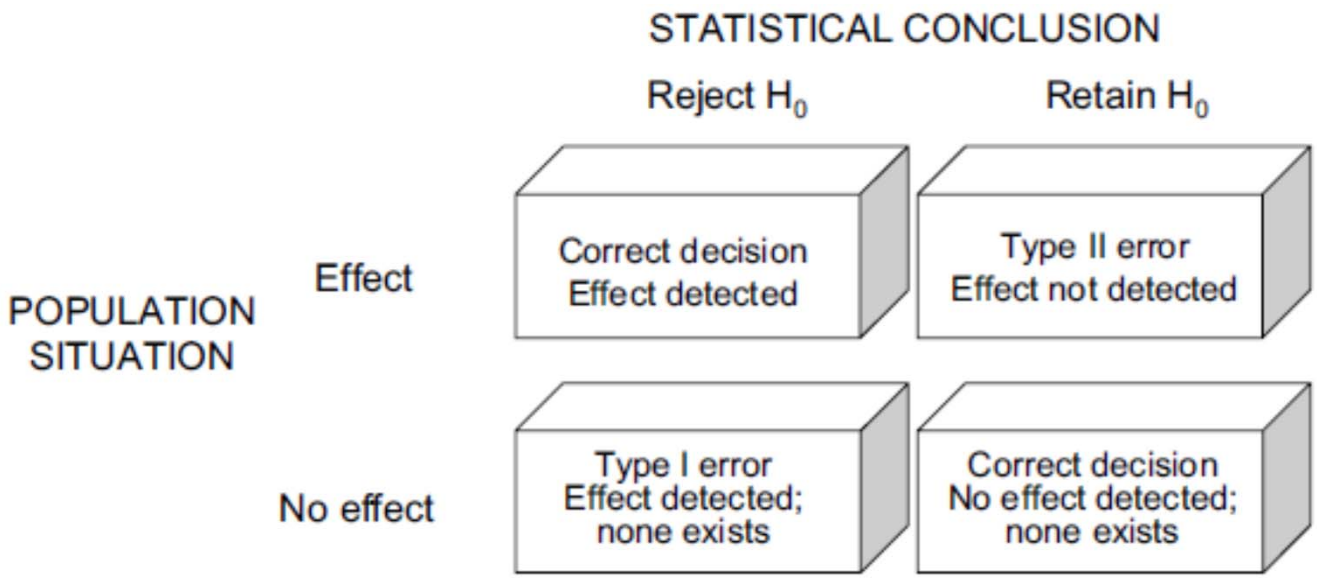
1. Evitar artefactos experimentales
2. Eliminar sesgos
   - Definiendo grupo control
   - Aleatorización
   - Diseño ciego
3. **Reducir el error de muestreo**
   - Independencia estadística de las réplicas
   - **Bloques y factores**



$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk}$$

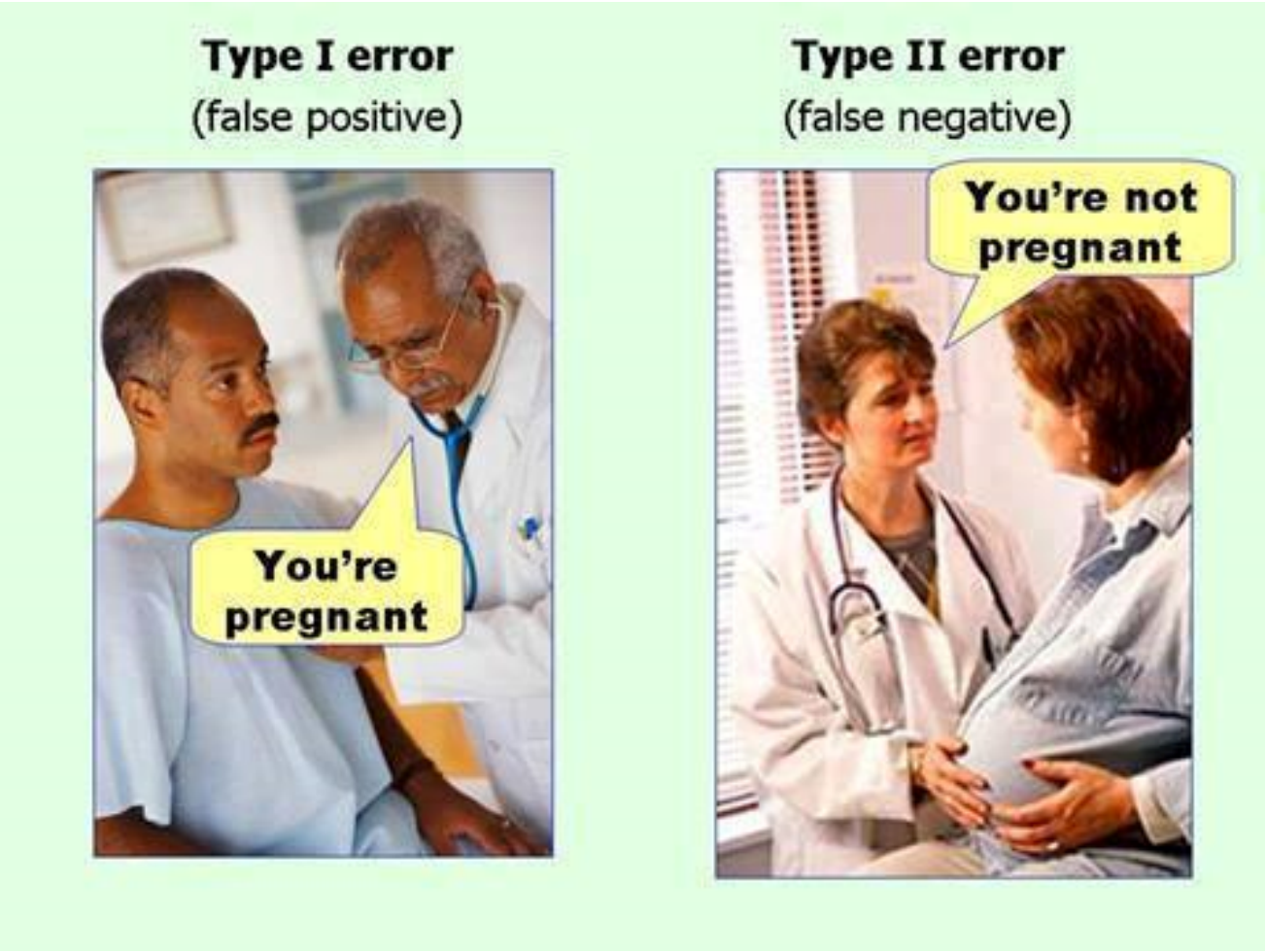
1. Evitar artefactos experimentales
2. Eliminar sesgos
   - Definiendo grupo control
   - Aleatorización
   - Diseño ciego
3. Reducir el error de muestreo
   - Independencia estadística de las réplicas
   - Bloques y factores
   - **Balance**

- Aumenta la potencia y robustez de los tests y permite estimas de varianza en efectos aleatorios. Prevenir el desgaste (reducción) de la muestra



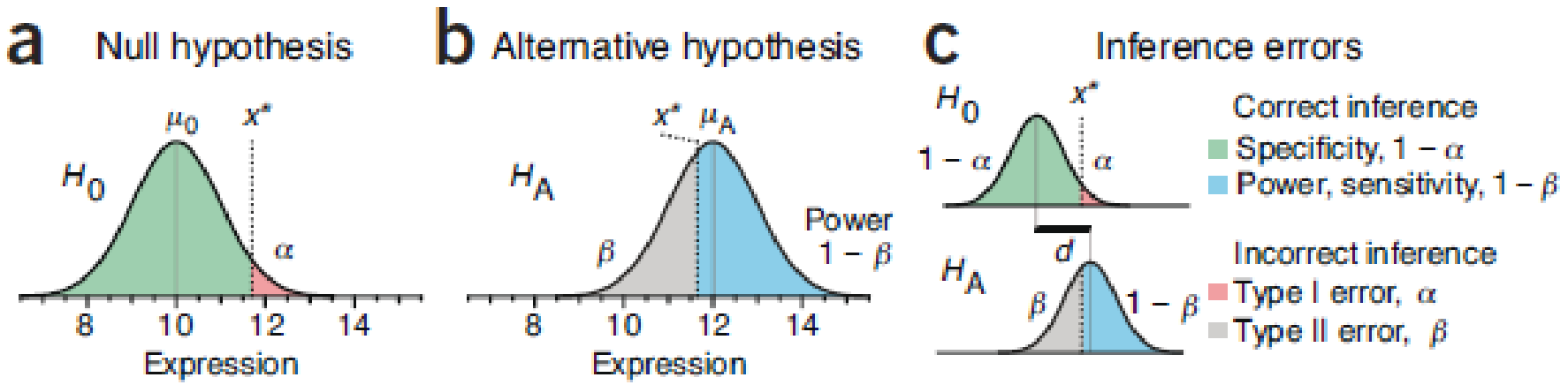
Error tipo I: Falso positivo (detecto un efecto que no existe) α
Error tipo II: Falso negativo (paso por alto un efecto que existe) β

Potencia: Probabilidad de que la muestra revele un efecto que existe 1-β

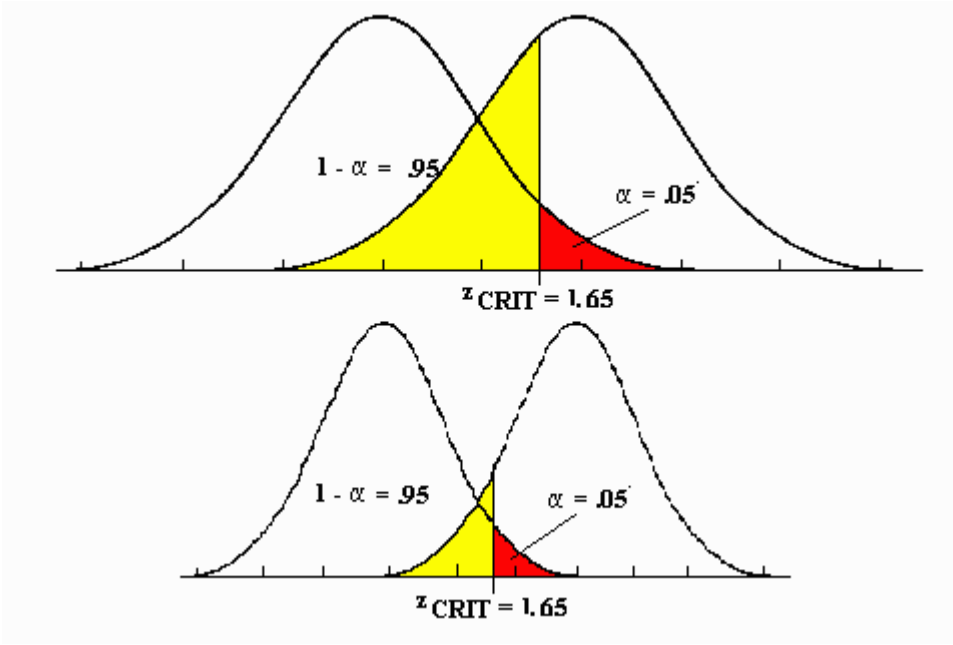


Error tipo I: Falso positivo (detecto un efecto que no existe) α
Error tipo II: Falso negativo (paso por alto un efecto que existe) β

Potencia: Probabilidad de que la muestra revele un efecto que existe 1-β

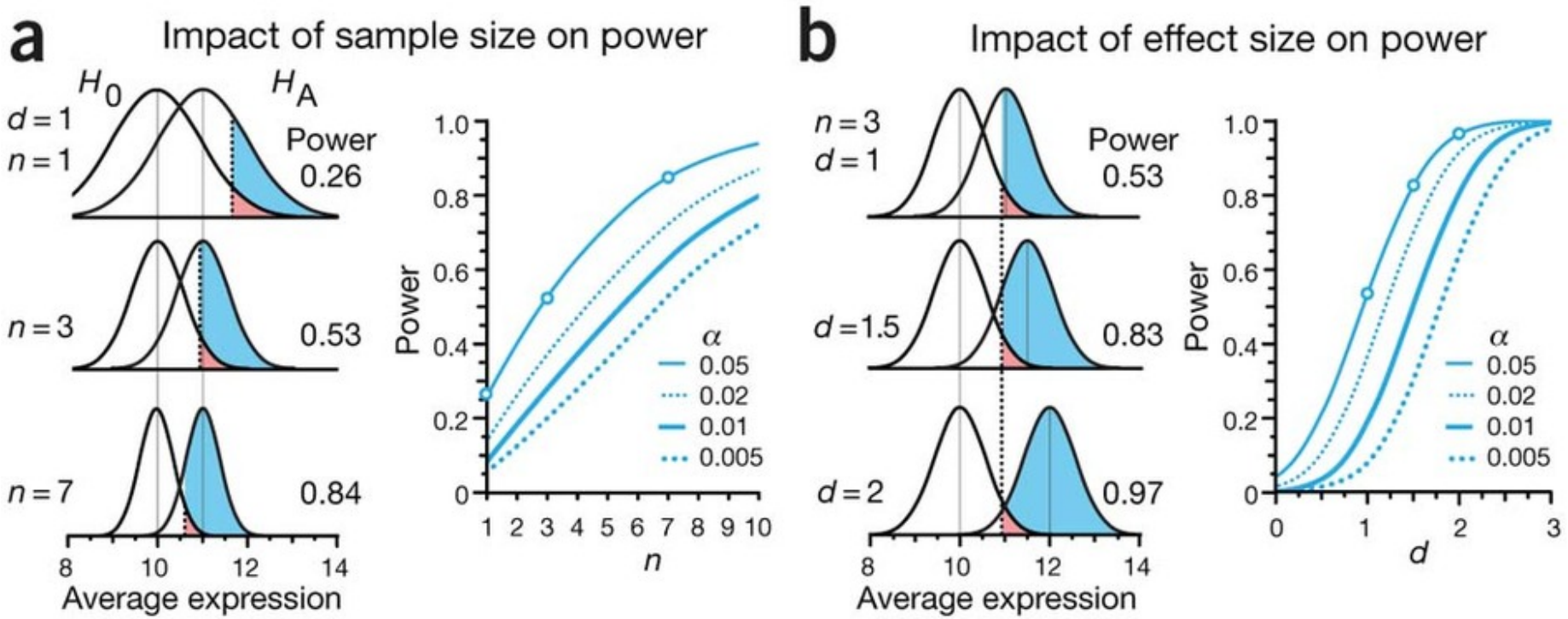


Las hembras en promedio pesan 10 y los machos pesan 12
¿Qué hago con un@ que pesa 11.6 o más?
Si decido que todo lo que pesa 11.6 o más es un macho y menos es una hembra:
- me puedo equivocar porque hay hembras mayores de 11.6 (falso positivo)
- pero también porque hay machos menores de 11. 6 (falso negativo)



Cuando yo tengo dos muestras para comparar con un test de hipótesis basado en hipótesis nula, los errores 1 y 2 son una función del grado de solapamiento de las muestras.

Las dos muestras son diferentes si las medias están muy separadas y/o las colas (errores típicos) son pequeñas

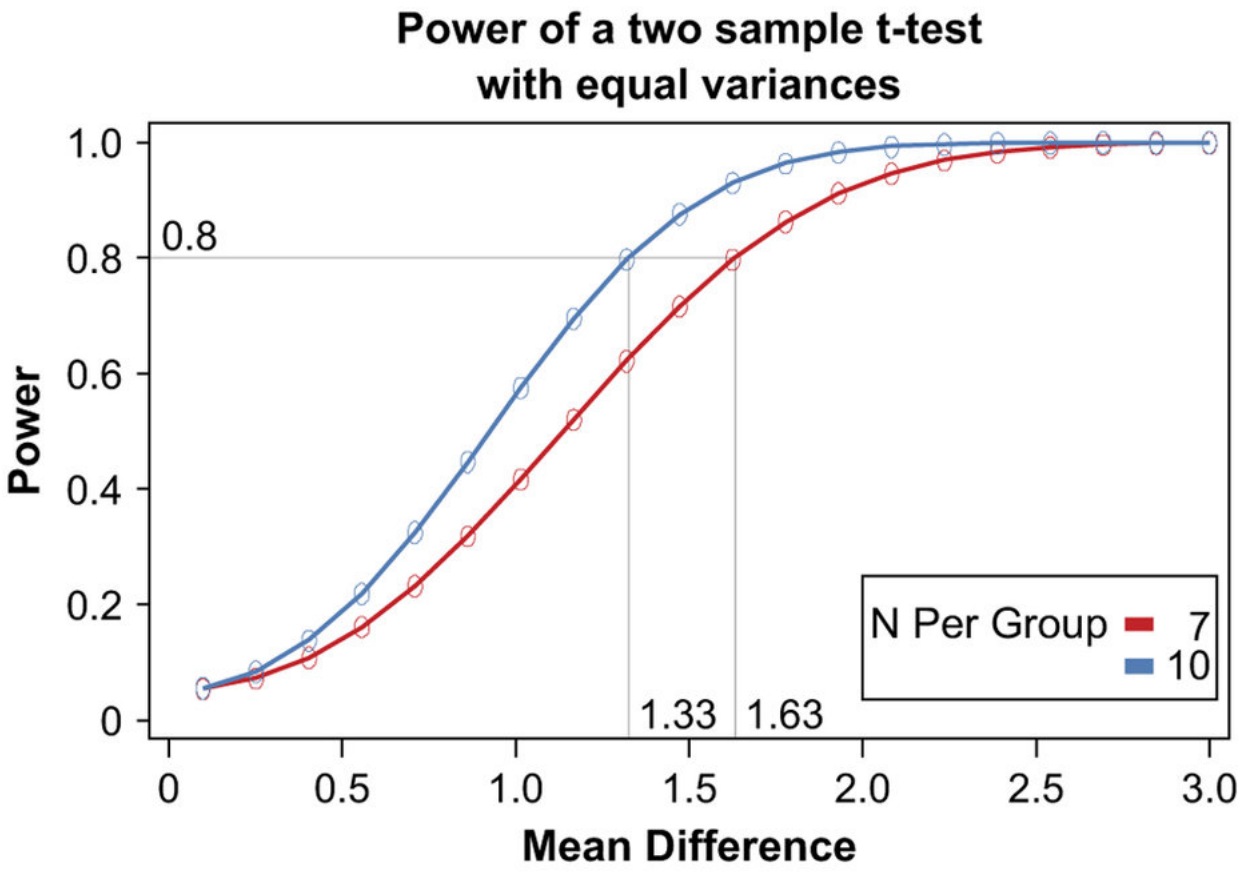


Cometería menos errores (de los dos tipos) si:

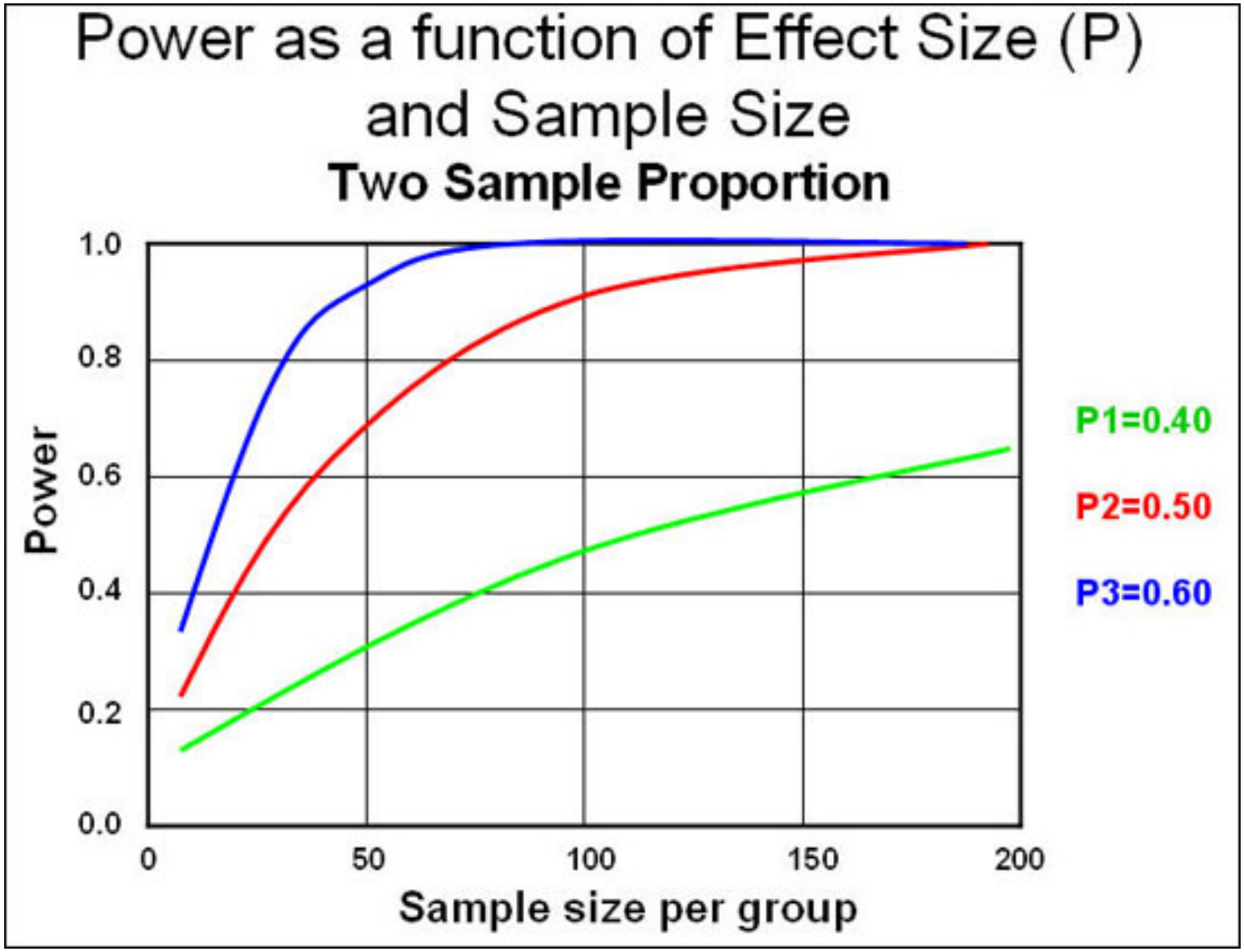
-Mi estima de los valores medios de machos y hembras tuviese poco error (estima basada en un amplio tamaño de muestra)

- Las diferencias de peso entre machos y hembras fuesen muy grandes (el tamaño del efecto es muy grande)

$$\text{Power} \propto \frac{ES\ \alpha\ \sqrt{n}}{\sigma}$$



Absolute standardized mean difference represented
alpha = 5%, two-tailed


Power as a function of Effect Size (P)
and Sample Size
Two Sample Proportion
Power
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
Sample size per group
P1=0.40
P2=0.50
P3=0.60

Análisis de potencia retrospectivo

- **No debe usarse para la interpretación de resultados NS sobre potencia "observada"**
- **Pero es útil si el ES estaba predefinido y en estudios comparativos (meta-análisis)**

acta ethol (2004) 7: 103–108
DOI 10.1007/s10211-004-0095-z

COMMENTARY

Shinichi Nakagawa · T. Mary Foster

The case against retrospective statistical power analyses with an introduction to power analysis

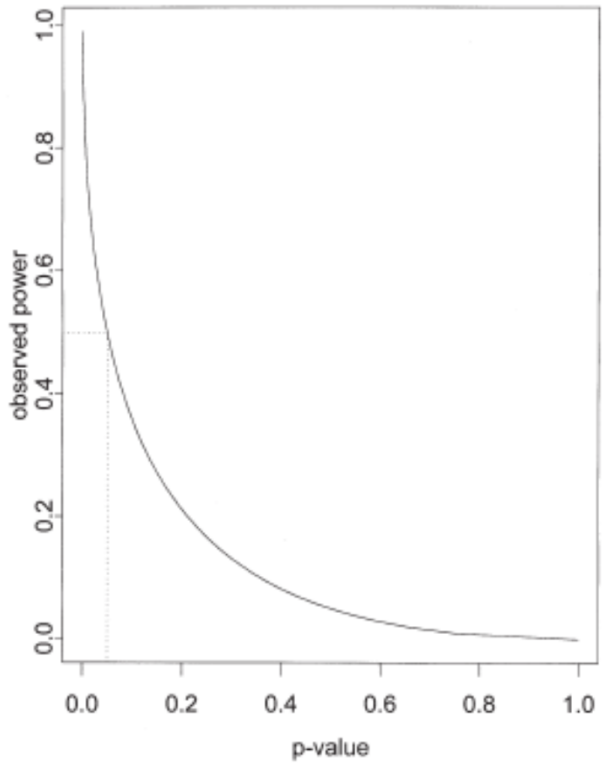


Figure 1. "Observed" Power as a Function of the p Value for a One-Tailed Z Test in Which α is Set to .05. When a test is marginally significant (P = .05) the estimated power is 50%.

- **Los resultados NS deben interpretarse en términos de Tamaño de Efecto o Ensayos de Equivalencia**

- **Un paper y un paquete de R para Ensayos de Equivalencia**

Understanding Equivalence and Noninferiority Testing

Esteban Walker, PhD and Amy S. Nowacki, PhD

The equivalence Package

September 7, 2006

Type Package

Title Provides tests and graphics for assessing tests of equivalence

Version 0.4.1

Date 2006-09-07

Author Andrew Robinson <A.Robinson@ms.unimelb.edu.au>

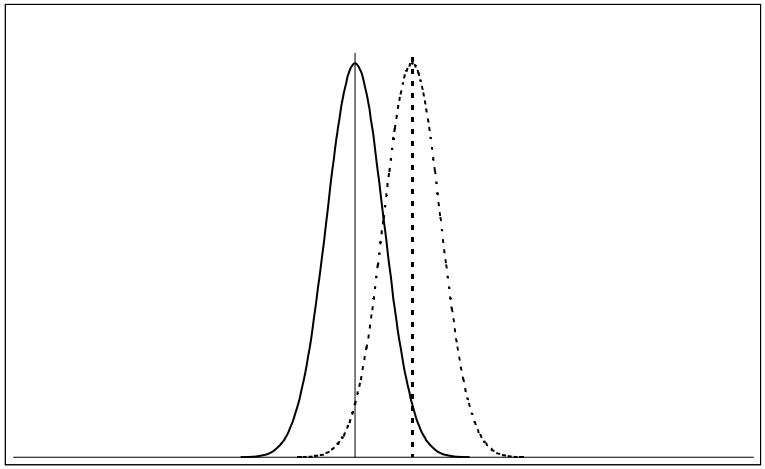
Maintainer Andrew Robinson <A.Robinson@ms.unimelb.edu.au>

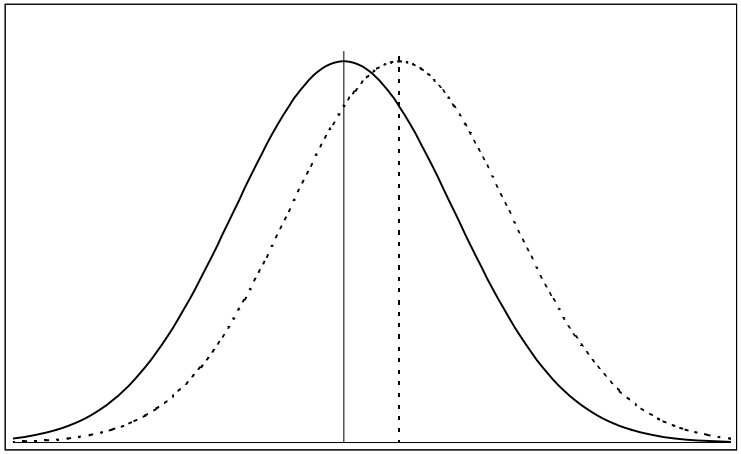
Análisis de potencia “a priori” (determinación de tamaño de muestra)

1. **Establecer claramente Ho y H1**
2. **Establecer el Modelo que usaremos para el análisis**
3. **Establecer α (0.05) y 1-β (0.8)**
4. **Verificar supuestos del Modelo (p. ej. datos transformados)**
5. **Obtener una estima de σ teniendo en cuenta la estructura de efectos (fijos y aleatorios) del Modelo (p. ej. un piloto)**
6. **Especificar el ES (p. ej. un piloto, literatura)**

$$Potencia \propto \frac{ES\ \alpha\ \sqrt{n}}{\sigma} \qquad Potencia \propto \frac{ES}{\sigma}\ \alpha\ \sqrt{n}$$

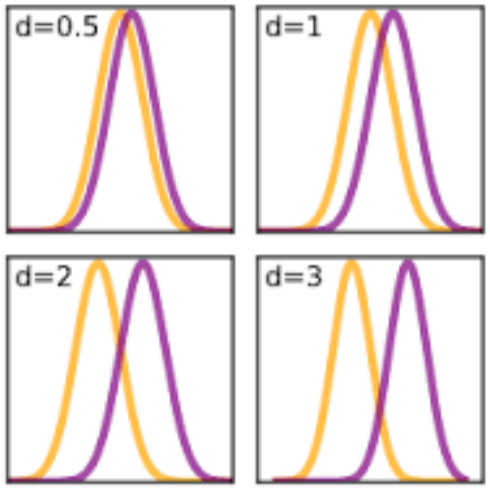
Tamaño de efecto estandarizado y no estandarizado





$$Potencia \propto \frac{ES\ \alpha\ \sqrt{n}}{\sigma} \qquad Potencia \propto \frac{ES}{\sigma}\ \alpha\ \sqrt{n}$$

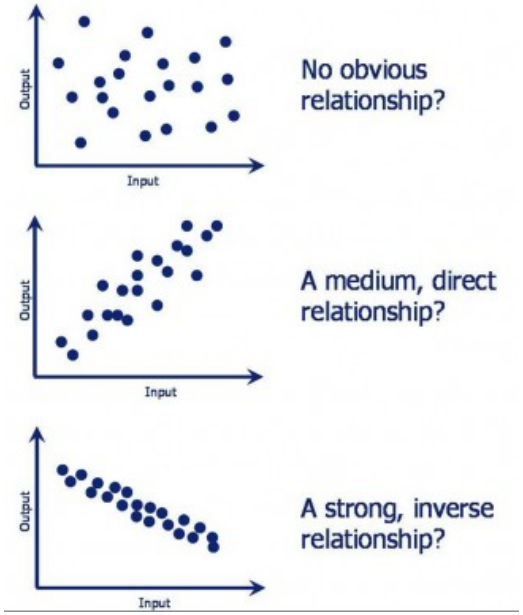
Los dos tipos de tamaño de efecto más usuales



$$d = \frac{\bar{x}_1 - \bar{x}_2}{s_{pooled}}$$

where

$$s_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$



$$\eta^2 = \frac{SS_{Effect}}{SS_{Total}}$$

$$\eta_P^2 = \frac{SS_{Effect}}{SS_{Effect} + SS_{Error}}$$

$$\omega^2 = \frac{SS_{Effect} - (df_{Effect})(MS_{Error})}{SS_{Total} + MS_{Error}}$$

Hay una relación entre los dos tipos de ES y a su vez con el estadístico t de Student

Group Difference Indices	Mean Contrast "Standardizer"
Cohen's 'd'	$\frac{M_1 - M_2}{S_{(1 \text{ or } 2)}}$
Hedge's 'g'	$\frac{M_1 - M_2}{S_{pooled}}$
Glass's 'Δ'	$\frac{M_{'t'} - M_{'c'}}{S_{'c'}}$
Relationship Index:	
r_{pb}	Point biserial correlation
Relationships:	
t and g	$t = g \times (\sqrt{(1/n_1 + 1/n_2)})^{-1}$
t and r_{pb}	$r_{pb} = t / (\sqrt{(t^2 + df_w)}$
r_{pb} and g	$r_{pb} = g / (\sqrt{(g^2 + df_w(1/n_1 + 1/n_2))}$

frontiers in PSYCHOLOGY

REVIEW ARTICLE
published: 26 November 2013
doi: 10.3389/fpsyg.2013.00863

Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for *t*-tests and ANOVAs

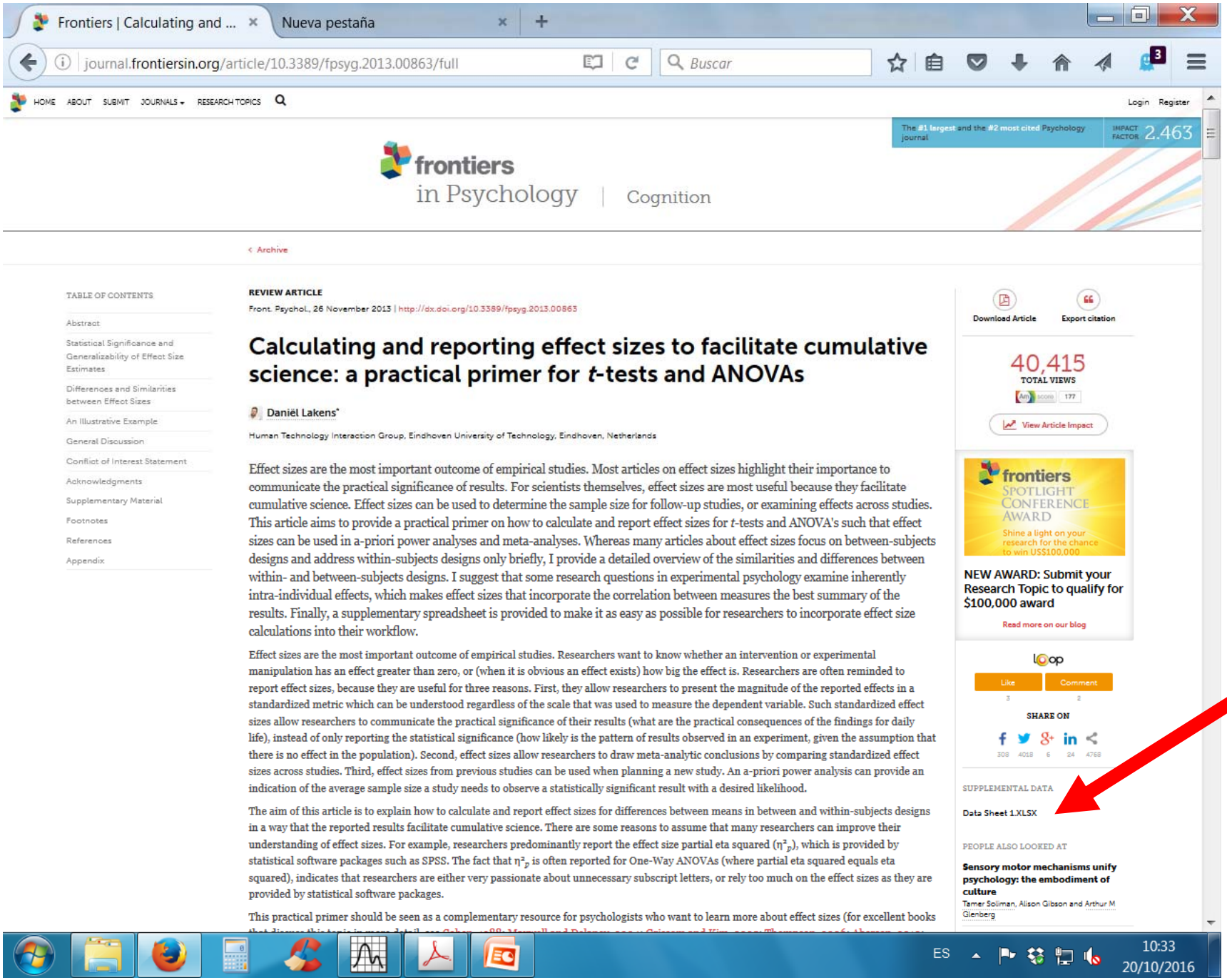
Daniël Lakens*

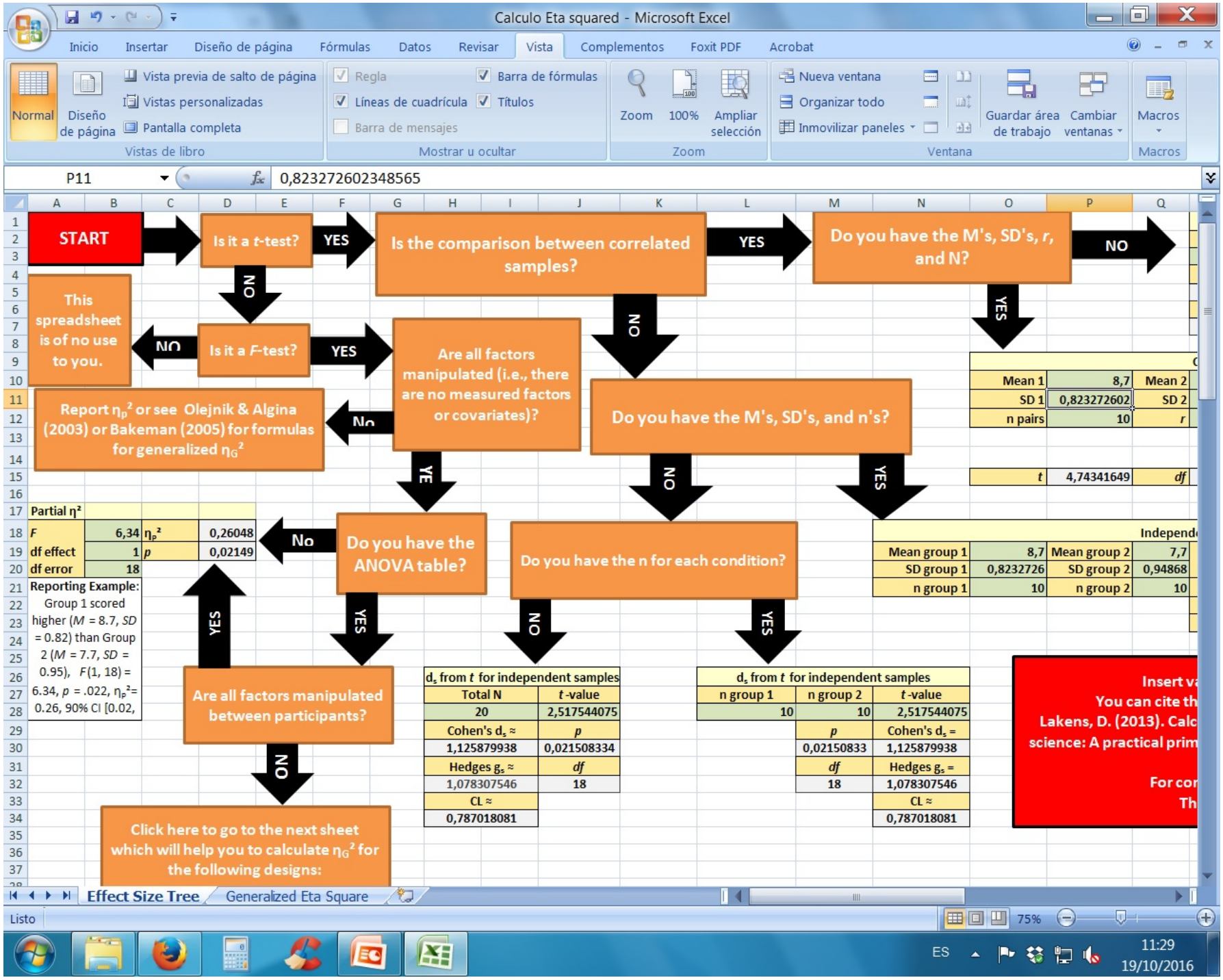
Table 1 | Summary of d family effect sizes, standardizers, and their recommended use.

ES	Standardizer	Use
Cohen's d_{pop}	σ (population)	Independent groups, use in power analyses when population σ is known, σ calculated with *n*
Cohen's d_s	Pooled *SD*	Independent groups, use in power analyses when population σ is unknown, σ calculated with *n*-1
Hedges' *g*	Pooled *SD*	Independent groups, corrects for bias in small samples, report for use in meta-analyses
Glass's Δ	*SD* pre measurement or control condition	Independent groups, use when experimental manipulation might affect the *SD*
Hedges' g_{av}	$(SD_1 + SD_2)/2$	Correlated groups, report for use in meta-analyses (generally recommended over Hedges' g_{rm})
Hedges' g_{rm}	*SD* difference scores corrected for correlation	Correlated groups, report for use in meta-analyses (more conservative then Hedges' g_{av})
Cohen's d_z	*SD* difference scores	Correlated groups, use in power analyses

Table 2 | Summary of *r* family effect sizes and their recommended use.

ES (Biased)	ES (Less Biased)	Use
eta squared (μ^2)	omega squared (ω^2)	Use for comparisons of effects within a single study
eta squared (μ_p^2)	omega squared (ω_p^2)	Use in power analyses, and for comparisons of effect sizes across studies with the same experimental design.
Generalized eta squared (μ_G^2)	Generalized omega squared (ω_G^2)	Use in meta-analyses to compare across experimental designs


Frontiers | Calculating and ...
Nueva pestaña
journal.frontiersin.org/article/10.3389/fpsyg.2013.00863/full
Buscar
HOME ABOUT SUBMIT JOURNALS RESEARCH TOPICS
Login Register
The #1 largest and the #2 most cited Psychology journal
IMPACT FACTOR 2.463
frontiers in Psychology | Cognition
< Archive
TABLE OF CONTENTS
Abstract
Statistical Significance and Generalizability of Effect Size Estimates
Differences and Similarities between Effect Sizes
An Illustrative Example
General Discussion
Conflict of Interest Statement
Acknowledgments
Supplementary Material
Footnotes
References
Appendix
REVIEW ARTICLE
Front. Psychol., 26 November 2013 | http://dx.doi.org/10.3389/fpsyg.2013.00863
Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs
Daniël Lakens*
Human Technology Interaction Group, Eindhoven University of Technology, Eindhoven, Netherlands
Effect sizes are the most important outcome of empirical studies. Most articles on effect sizes highlight their importance to communicate the practical significance of results. For scientists themselves, effect sizes are most useful because they facilitate cumulative science. Effect sizes can be used to determine the sample size for follow-up studies, or examining effects across studies. This article aims to provide a practical primer on how to calculate and report effect sizes for t-tests and ANOVA's such that effect sizes can be used in a-priori power analyses and meta-analyses. Whereas many articles about effect sizes focus on between-subjects designs and address within-subjects designs only briefly, I provide a detailed overview of the similarities and differences between within- and between-subjects designs. I suggest that some research questions in experimental psychology examine inherently intra-individual effects, which makes effect sizes that incorporate the correlation between measures the best summary of the results. Finally, a supplementary spreadsheet is provided to make it as easy as possible for researchers to incorporate effect size calculations into their workflow.
Effect sizes are the most important outcome of empirical studies. Researchers want to know whether an intervention or experimental manipulation has an effect greater than zero, or (when it is obvious an effect exists) how big the effect is. Researchers are often reminded to report effect sizes, because they are useful for three reasons. First, they allow researchers to present the magnitude of the reported effects in a standardized metric which can be understood regardless of the scale that was used to measure the dependent variable. Such standardized effect sizes allow researchers to communicate the practical significance of their results (what are the practical consequences of the findings for daily life), instead of only reporting the statistical significance (how likely is the pattern of results observed in an experiment, given the assumption that there is no effect in the population). Second, effect sizes allow researchers to draw meta-analytic conclusions by comparing standardized effect sizes across studies. Third, effect sizes from previous studies can be used when planning a new study. An a-priori power analysis can provide an indication of the average sample size a study needs to observe a statistically significant result with a desired likelihood.
The aim of this article is to explain how to calculate and report effect sizes for differences between means in between and within-subjects designs in a way that the reported results facilitate cumulative science. There are some reasons to assume that many researchers can improve their understanding of effect sizes. For example, researchers predominantly report the effect size partial eta squared (η^2_p), which is provided by statistical software packages such as SPSS. The fact that η^2_p is often reported for One-Way ANOVAs (where partial eta squared equals eta squared), indicates that researchers are either very passionate about unnecessary subscript letters, or rely too much on the effect sizes as they are provided by statistical software packages.
This practical primer should be seen as a complementary resource for psychologists who want to learn more about effect sizes (for excellent books
Download Article
Export citation
40,415
TOTAL VIEWS
View Article Impact
frontiers SPOTLIGHT CONFERENCE AWARD
Shine a light on your research for the chance to win US$100,000
NEW AWARD: Submit your Research Topic to qualify for $100,000 award
Read more on our blog
Like
Comment
SHARE ON
SUPPLEMENTAL DATA
Data Sheet 1.XLSX
PEOPLE ALSO LOOKED AT
Sensory motor mechanisms unify psychology: the embodiment of culture
Tamer Soliman, Alison Gibson and Arthur M Glenberg
ES
10:33
20/10/2016


Calculo Eta squared - Microsoft Excel
Inicio
Insertar
Diseño de página
Fórmulas
Datos
Revisar
Vista
Complementos
Foxit PDF
Acrobat
Normal
Diseño de página
Vista previa de salto de página
Vistas personalizadas
Pantalla completa
Vistas de libro
Regla
Líneas de cuadrícula
Barra de mensajes
Barra de fórmulas
Títulos
Mostrar u ocultar
Zoom
100%
Ampliar selección
Zoom
Nueva ventana
Organizar todo
Inmovilizar paneles
Guardar área de trabajo
Cambiar ventanas
Ventana
Macros
Macros
P11
0,823272602348565
START
Is it a t-test?
YES
Is the comparison between correlated samples?
YES
Do you have the M's, SD's, r, and N?
NO
YES
NO
This spreadsheet is of no use to you.
NO
Is it a F-test?
YES
Are all factors manipulated (i.e., there are no measured factors or covariates)?
NO
Do you have the M's, SD's, and n's?
Report ηp² or see Olejnik & Algina (2003) or Bakeman (2005) for formulas for generalized ηG²
NO
YES
NO
YES
Partial η²
F 6,34 ηp² 0,26048
df effect 1 p 0,02149
df error 18
Reporting Example: Group 1 scored higher (M = 8.7, SD = 0.82) than Group 2 (M = 7.7, SD = 0.95), F(1, 18) = 6.34, p = .022, ηp²= 0.26, 90% CI [0.02,
Do you have the ANOVA table?
NO
YES
Do you have the n for each condition?
NO
YES
Are all factors manipulated between participants?
YES
NO
Click here to go to the next sheet which will help you to calculate ηG² for the following designs:
Mean 1 8,7 Mean 2
SD 1 0,823272602 SD 2
n pairs 10 r
t 4,74341649 df
Mean group 1 8,7 Mean group 2 7,7
SD group 1 0,8232726 SD group 2 0,94868
n group 1 10 n group 2 10
ds from t for independent samples
Total N t-value
20 2,517544075
Cohen's ds ≈ p
1,125879938 0,021508334
Hedges gs ≈ df
1,078307546 18
CL ≈
0,787018081
ds from t for independent samples
n group 1 n group 2 t-value
10 10 2,517544075
p Cohen's ds =
0,02150833 1,125879938
df Hedges gs =
18 1,078307546
CL ≈
0,787018081
Insert v
You can cite th
Lakens, D. (2013). Calc
science: A practical prim
For cor
Th
Effect Size Tree
Generalized Eta Square
Listo
75%
ES
11:29
19/10/2016

Una web más simple que el trabajo anterior donde calcular tamaños de efecto

https://www.psychometrica.de/effect_size.html

1. Comparison of groups with equal size (*Cohen's d, Glass Δ*)
2. Comparison of groups with different sample size(*Cohen's d, Hedges' g*)
3. Effect size for pre-post-control studies with the correction of pretest differences
4. Calculation of d from the test statistics of dependent and independent t-tests
5. Computation of d from the F-value of Analyses of Variance (ANOVA)
6. Calculation of effect sizes from ANOVAs with multiple groups, based on group means
7. Increase of success through intervention: The Binomial Effect Size Display (BESD) and Number Needed to Treat (NNT)
8. Risk Ratio, Odds Ratio and Risk Difference
9. Effect size for the difference between two correlations
10. Effect size calculator for non-parametric Tests: Mann-Whitney-U, Wilcoxon-W and Kruskal-Wallis-H
11. Computation of the pooled standard deviation
12. Transformation of the effect sizes *r, d, f, Odds Ratio*and *eta square*
13. Computation of the effect sizes *d*, *r* and η^2 from χ^2- and *z* test statistics
14. Table for interpreting the magnitude of d, r and eta square according to Hattie (2009) and Cohen (1988)

Siempre es posible calcular el ES de un modelo (mixto) a mano con tal de que entendamos bien los grados de libertad

Biol. Rev. (2007), **82**, pp. 591–605.
doi:10.1111/j.1469-185X.2007.00027.x

Effect size, confidence interval and statistical significance: a practical guide for biologists

Shinichi Nakagawa[1,*] and Innes C. Cuthill[2]

$$d = \frac{t(n_1 + n_2)}{\sqrt{n_1 n_2}\sqrt{df}},$$

$$r = \frac{t}{\sqrt{t^2 + df}}$$

¿Qué modelo? El que elijamos (preferiblemente con una selección basada en AIC o BIC)

O en último término recurrir a los ES (pequeño, mediano, grande) de Cohen. Cualquier ES se corresponde con un tamaño de muestra para un test determinado

Group size as a function of S/N ratio (t test, 5% sig., 2-sided)

SN ratio	90% power	80% power
0.2	526	393
0.4	132	99
0.6	59	45
0.8	34	26
1.0	22	17
1.2	16	12
1.4	12	9
1.6	9	7
1.8	8	6
2.0	6	5
2.2	6	4
2.4	5	4
2.6	4	4
2.8	4	3
3.0	4	3

Sample size in each group for comparing two proportions
(power=0.8, significance level=0.05)

	Percent for group 1									
% Group 2	0	10	20	30	40	50	60	70	80	90
10	74									
20	34	199								
30	21	62	293							
40	15	32	81	356						
50	11	20	39	93	387					
60	8	13	23	42	97	387				
70	6	10	14	23	42	93	356			
80	5	7	10	15	23	39	81	293		
90	4	5	7	10	14	20	32	62	199	
100	2	4	5	6	8	11	15	21	34	74

Cómo interpretar un ES

Effect Size	Percentage of control group who would be below average person in experimental group	Rank of person in a control group of 25 who would be equivalent to the average person in experimental group	Probability that you could guess which group a person was in from knowledge of their 'score'.	Equivalent correlation, *r* (=Difference in percentage 'successful' in each of the two groups, BESD)	Probability that person from experimental group will be higher than person from control, if both chosen at random (=CLES)
0.0	50%	13th	0.50	0.00	0.50
0.1	54%	12th	0.52	0.05	0.53
0.2	58%	11th	0.54	0.10	0.56
0.3	62%	10th	0.56	0.15	0.58
0.4	66%	9th	0.58	0.20	0.61
0.5	69%	8th	0.60	0.24	0.64
0.6	73%	7th	0.62	0.29	0.66
0.7	76%	6th	0.64	0.33	0.69
0.8	79%	6th	0.66	0.37	0.71
0.9	82%	5th	0.67	0.41	0.74
1.0	84%	4th	0.69	0.45	0.76
1.2	88%	3rd	0.73	0.51	0.80
1.4	92%	2nd	0.76	0.57	0.84
1.6	95%	1st	0.79	0.62	0.87
1.8	96%	1st	0.82	0.67	0.90
2.0	98%	1st (or 1st out of 44)	0.84	0.71	0.92
2.5	99%	1st (or 1st out of 160)	0.89	0.78	0.96
3.0	99.9%	1st (or 1st out of 740)	0.93	0.83	0.98

Distintas marcas (indices) de ES tienen distintas tallas de camisetas

Test	ES Index	Small	Medium	Large
Comparison of independent means	Cohen's *d*	0.20	0.50	0.80
Correlation	*r*	0.10	0.30	0.50
P=0.5 and the sign test	*g= P-0.5*	0.05	0.15	0.25
Difference between proportions.	*h = arcsin(P1)-arcsin(P2)*	0.20	0.50	0.80
Crosstabulation, chi-square for goodness of fit.	*w*	0.10	0.30	0.50
ANOVA, one-way	*f*	0.10	0.25	0.40
Multiple regresion	f^2	0.02	0.15	0.35

Dos programas para análisis de potencia

http://biomath.info/power/

Behavior Research Methods
2007, 39 (2), 175-191

G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences

FRANZ FAUL
Christian-Albrechts-Universität Kiel, Kiel, Germany

EDGAR ERDFELDER
Universität Mannheim, Mannheim, Germany

AND

ALBERT-GEORG LANG AND AXEL BUCHNER
Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

Behavior Research Methods
2009, 41 (4), 1149-1160
doi:10.3758/BRM.41.4.1149

Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses

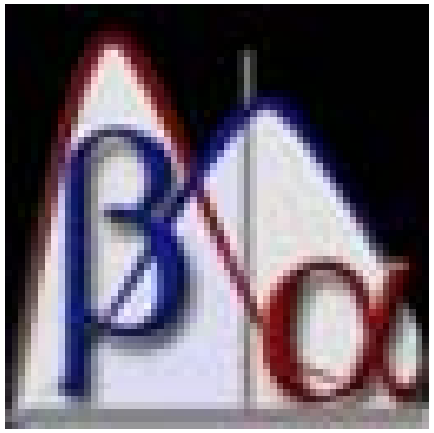
FRANZ FAUL
Christian-Albrechts-Universität, Kiel, Germany

EDGAR ERDFELDER
Universität Mannheim, Mannheim, Germany

AND

AXEL BUCHNER AND ALBERT-GEORG LANG
Heinrich-Heine-Universität, Düsseldorf, Germany

En G*Power 3 el tamaño de efecto para una prueba F entra en forma de *f*



Behavior Research Methods
2007, 39 (2), 175-191

G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences

FRANZ FAUL
Christian-Albrechts-Universität Kiel, Kiel, Germany

EDGAR ERDFELDER
Universität Mannheim, Mannheim, Germany

AND

ALBERT-GEORG LANG AND AXEL BUCHNER
Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

groups. The effect size index f is defined as:

$$f = \sqrt{\frac{\eta^2}{1-\eta^2}}, \tag{7}$$

with η^2 as the amount of the total population variance explained by the group differences specified in H_1. In case of unequal group sample sizes n_j, the effect size index f is calculated as follows:

$$f = \frac{\sqrt{\frac{\sum_{j=1}^{k} n_j \cdot (\mu_j - \bar{\mu})^2}{N}}}{\sigma}. \tag{8}$$

In Equation (8), n_j denotes the number of subjects, μ_j the population mean of group j, $\bar{\mu} = (\sum_{j=1}^{k} n_j \cdot \mu_j)/N$ the weighted mean of the k population means, N the total sample size, and σ the population standard deviation in each group.

Un paquete de R para análisis de potencia y simulaciones de modelos mixtos

Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2016

doi: 10.1111/2041-210X.12504

APPLICATION

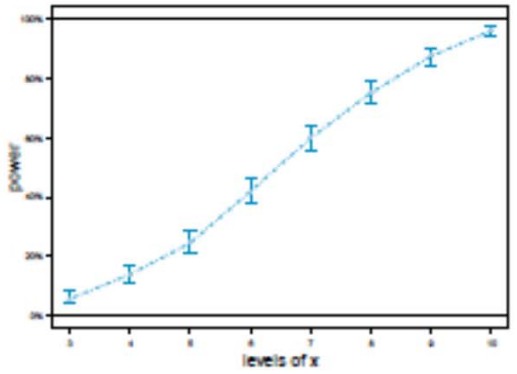
SIMR: an R package for power analysis of generalized linear mixed models by simulation

Peter Green* and Catriona J. MacLeod

Landcare Research, Private Bag 1930, Dunedin 9054, New Zealand

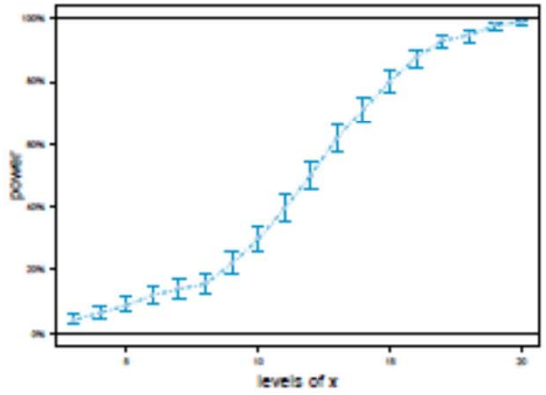
With `simr`, these steps are automated and a power curve can be calculated and plotted with just a few lines of code:

```
library(simr)
fit <- lmer(y ~ x + (1|g),
            data=example)
pa <- power(fit)
plot(pa)
```



This analysis shows that we have insufficient power for any of the sample sizes considered. We can extend the maximum number of levels of x in the analysis:

```
fit <- extend(fit, along='x', n=20)
plot(power(fit))
```



Un paquete (carísimo) de análisis de potencia con buenos manuales sobre el tema (gratis)

https://www.ncss.com/software/pass/

PASS Documentation

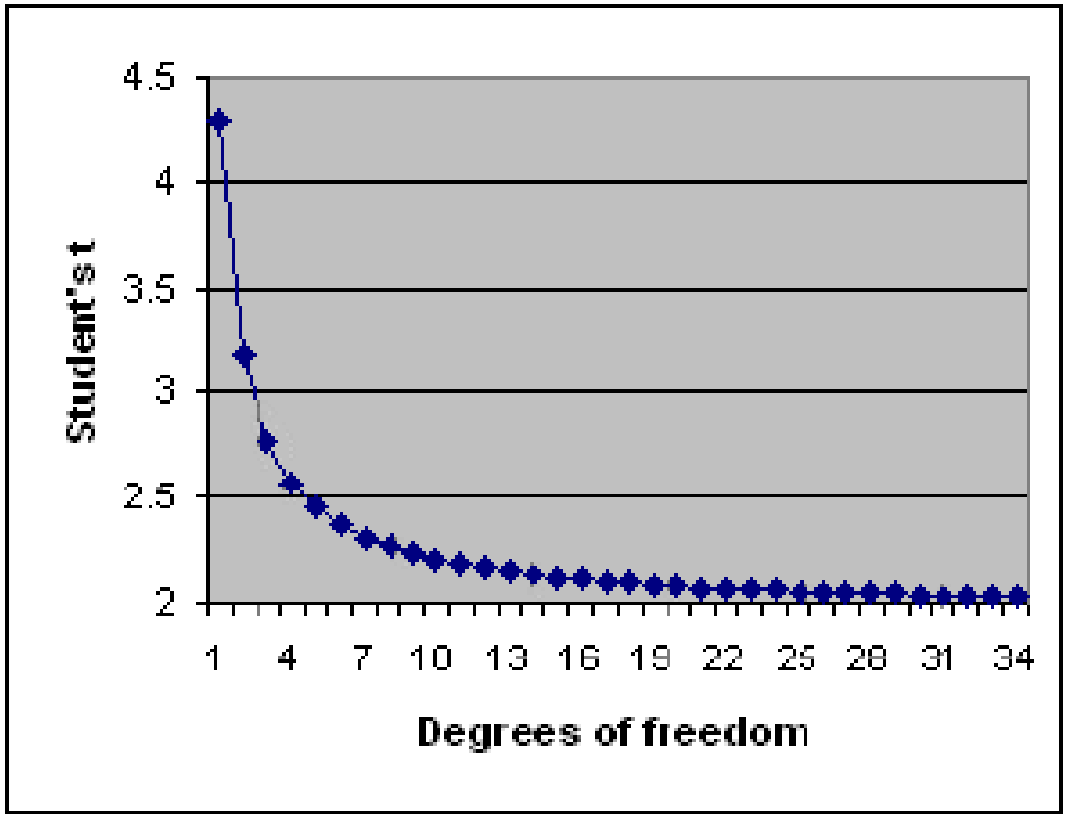
Use the links below to load individual chapters from the PASS statistical software training documentation in PDF format. The chapters correspond to the procedures available in PASS. Each chapter generally has an introduction to the topic, technical details including power and sample size calculation details, explanations for the procedure options, examples, and procedure validation examples. Each of these chapters is also available through the PASS help system when running the software.

Jump to topic:

El último recurso cuando no hay forma de que podamos recurrir a un análisis de potencia

Ecuación de recursos

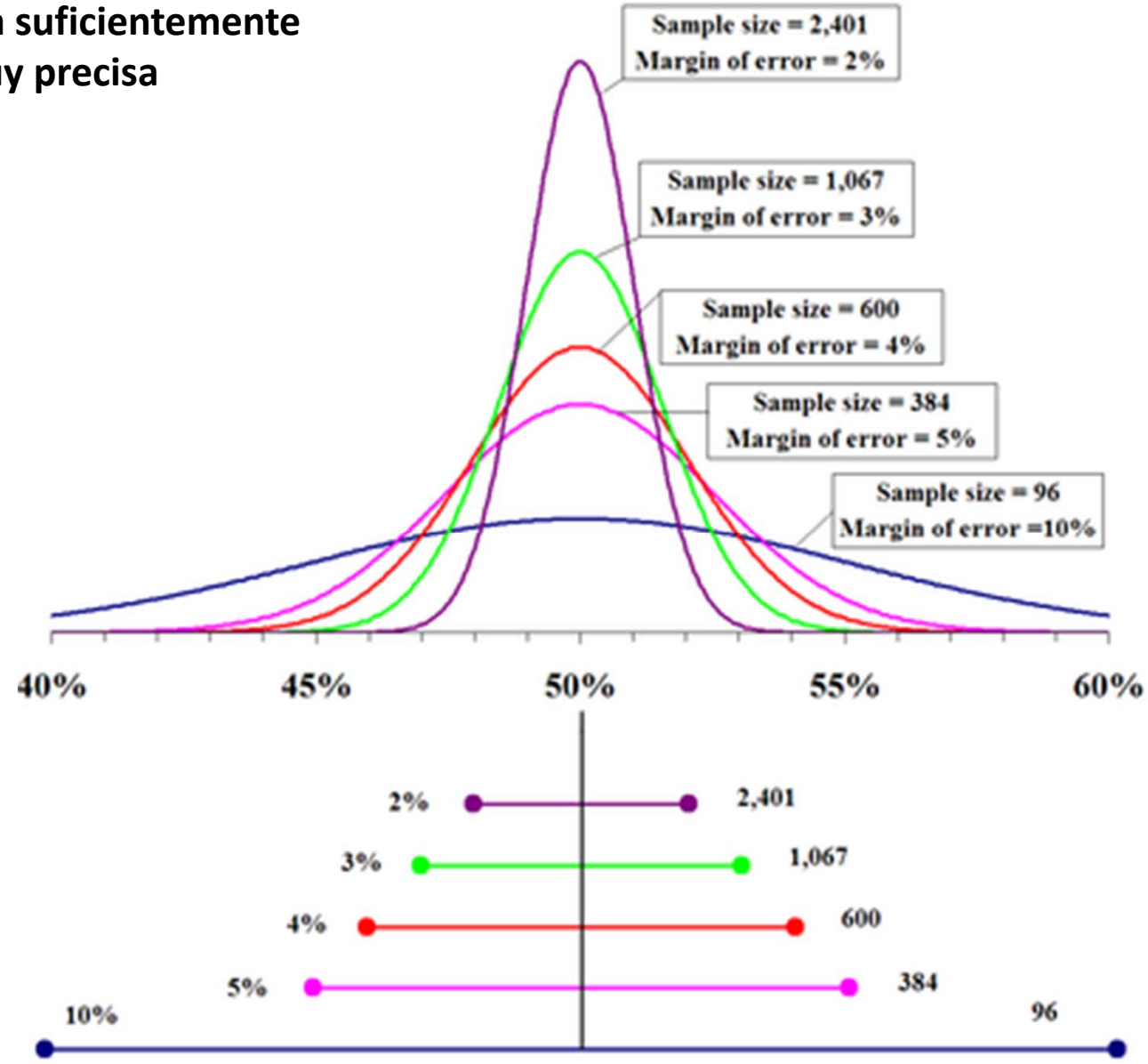
$E = N - B - T - 1$; $10 < E < 20$



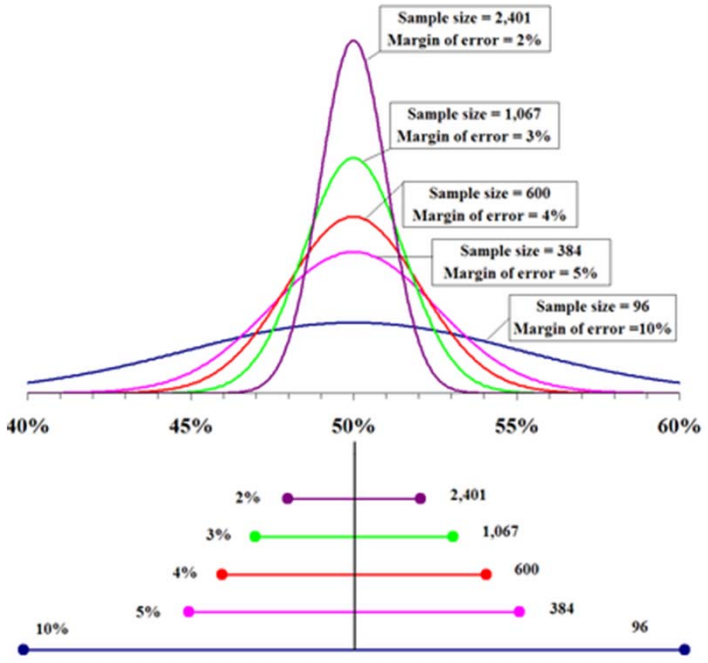
Mead R. 1988. The design of experiments. Cambridge, New York: Cambridge Univ. Press

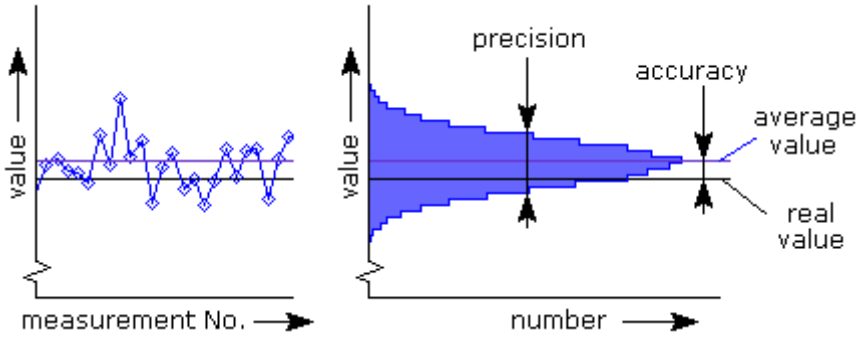
Crítica del paradigma de Ensayos de Hipótesis Basados en Hipótesis Nulas

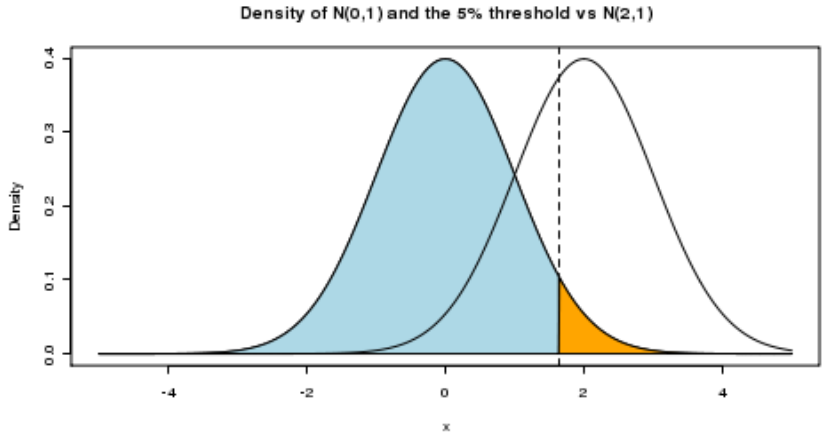
Toda muestra suficientemente grande es muy precisa

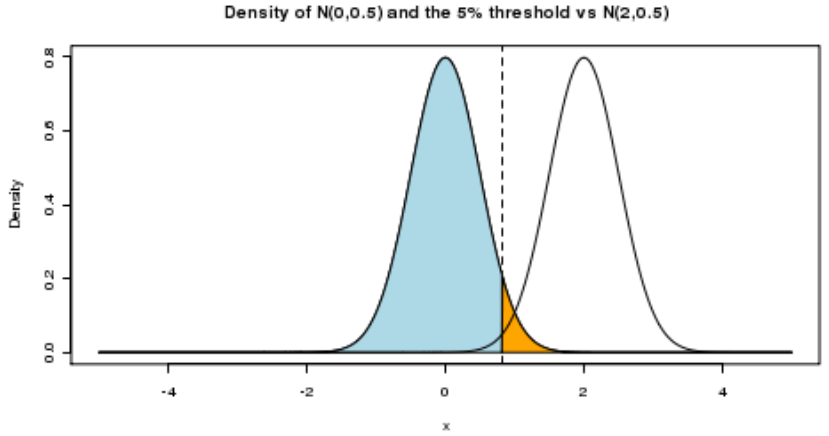


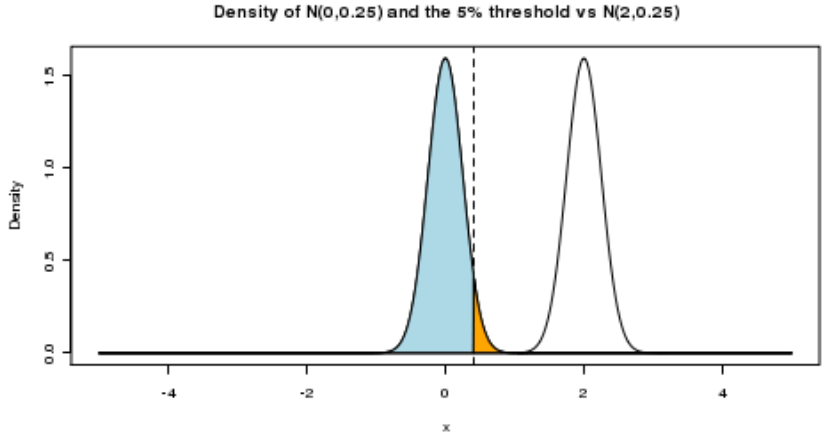
Ho nunca es cierta, en realidad



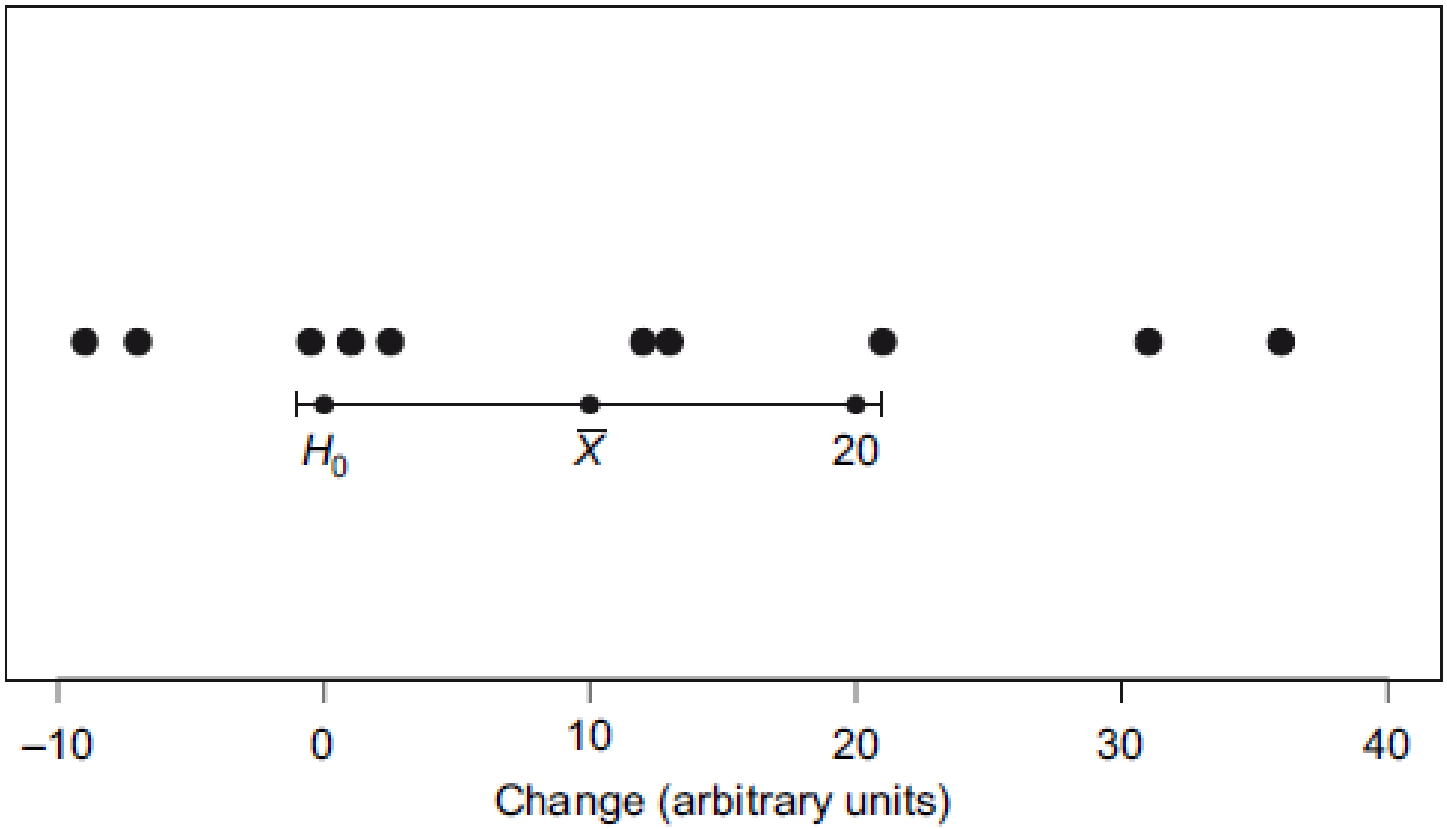




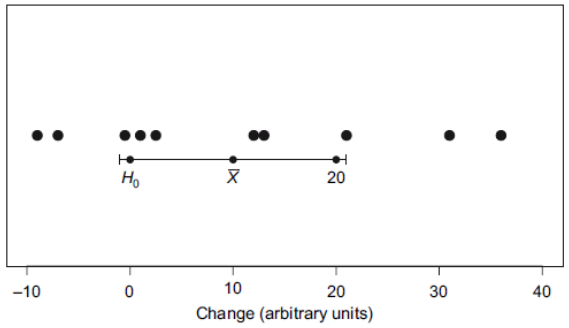


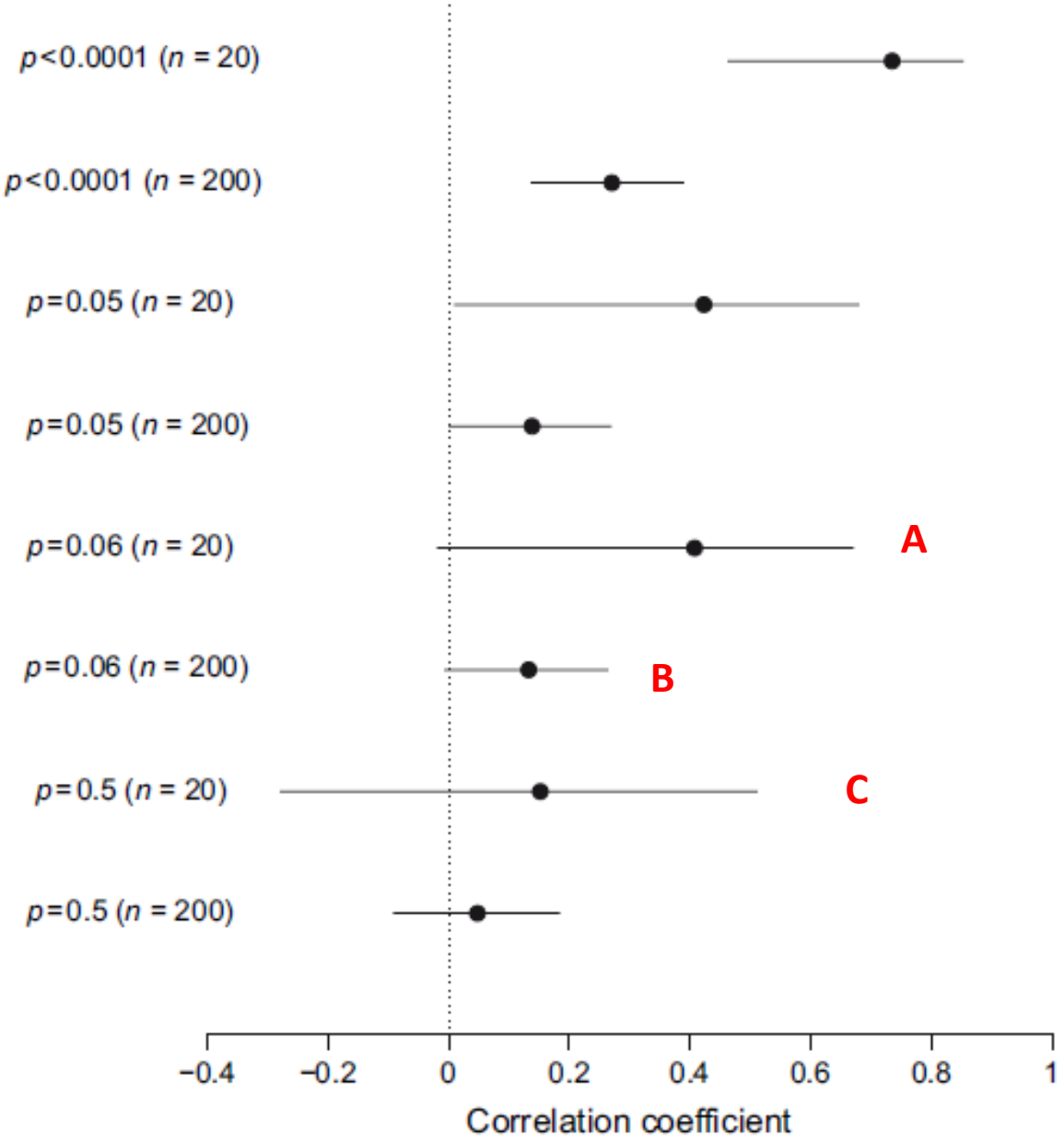


Un tamaño de efecto nulo (equivalente a rechazar Ho) es tan probable como un tamaño de efecto 20 en un experimento futuro y esto lo ignora el EHBHN



En el paradigma del EHBHN, los experimentos A, B y C son todos igual (de malos)

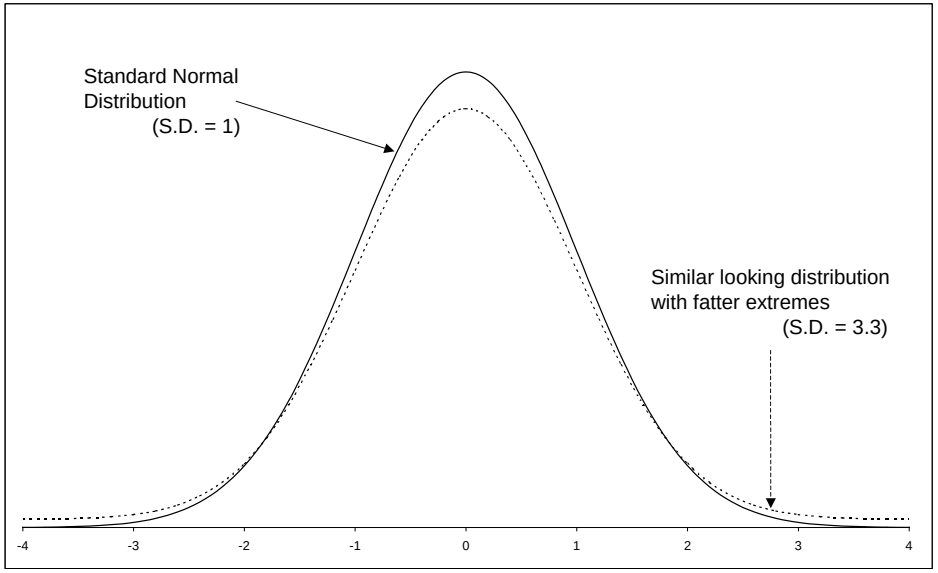


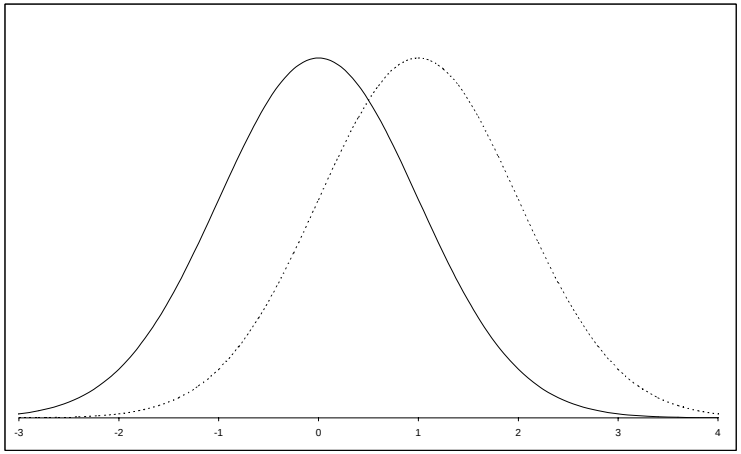


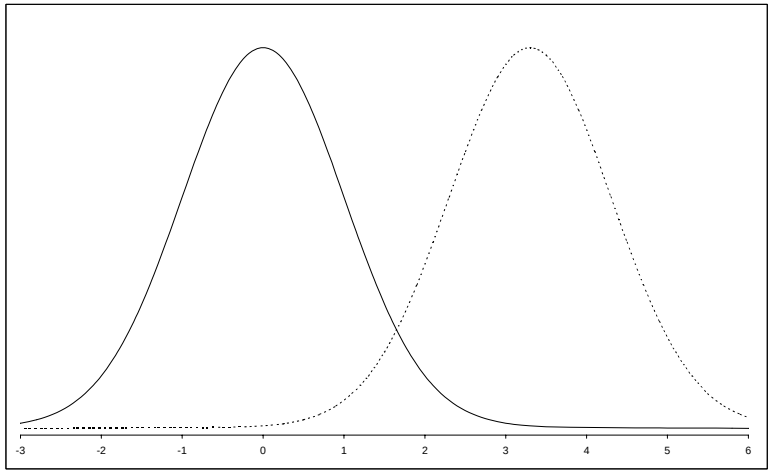
El cálculo de IC de tamaños de efecto basado en errores típicos es muy sensible a violaciones del supuesto de normalidad

$$\sigma[d] = \sqrt{\frac{N_E + N_C}{N_E \times N_C} + \frac{d^2}{2(N_E + N_C)}}$$

95 % CI = d – 1.96 × $\sigma[d]$ *to* **d + 1.96 × $\sigma[d]$**



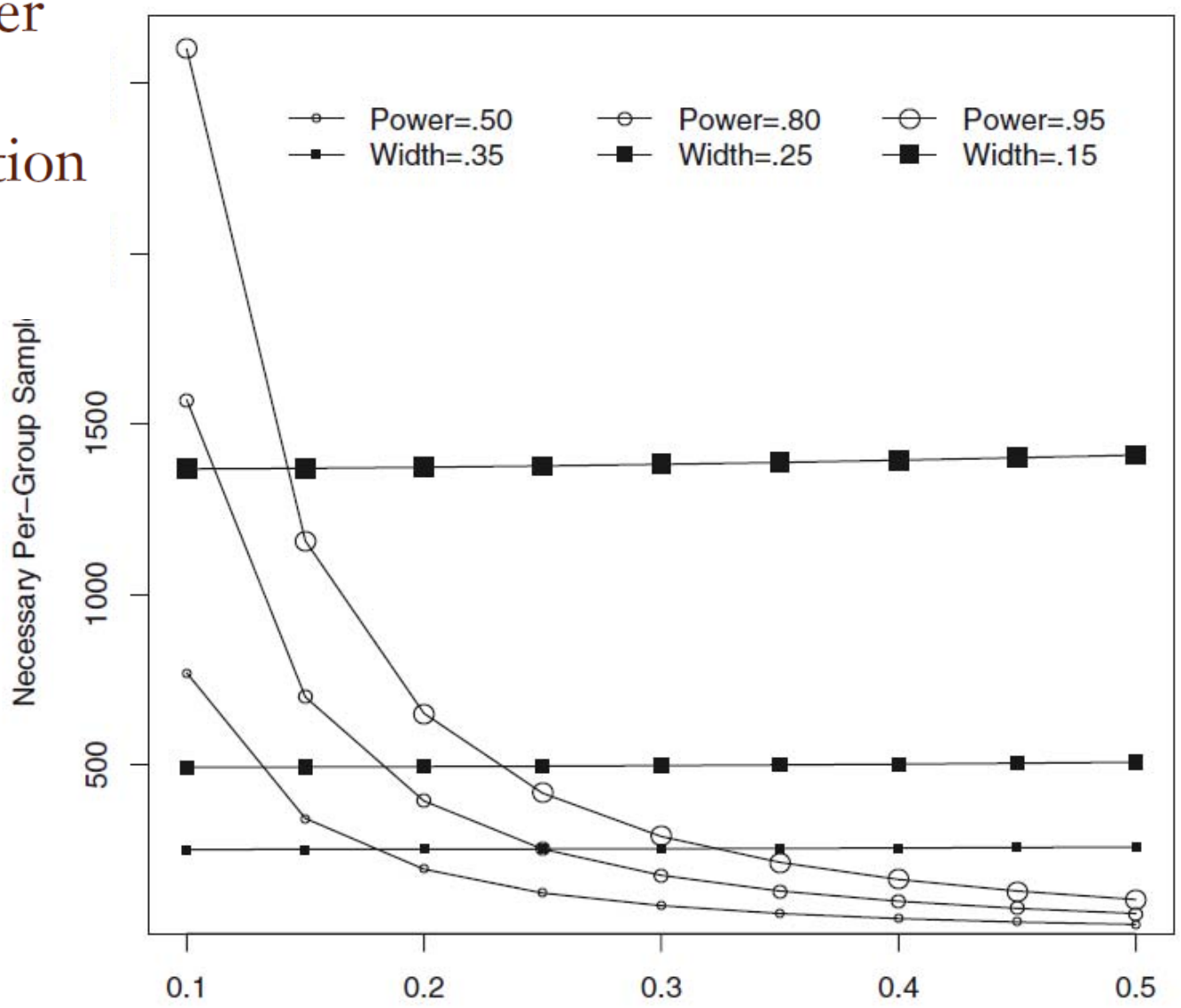




Determinación de tamaño de muestra en análisis de exactitud AIPE

Sample Size Planning for Statistical Power and Accuracy in Parameter Estimation

Scott E. Maxwell,[1] Ken Kelley,[2] and Joseph R. Rausch[3]



Un paquete de R para cálculo de IC para ES y análisis AIPE basados en distribuciones no centradas

Behavior Research Methods
2007, 39 (4), 979-984

Methods for the Behavioral, Educational, and Social Sciences: An R package

KEN KELLEY
Indiana University, Bloomington, Indiana

Package 'MBESS'

September 23, 2016

Type Package

Title The MBESS R Package

Version 4.1.0

Date 2016-09-21

Maintainer Ken Kelley <kkelley@nd.edu>

Depends R (>= 3.2.0), stats

For example, a confidence interval for the standardized mean difference can be obtained with the `ci.smd()` function. An example call to the `ci.smd()` function could be of the form

```
R> ci.smd(smd=d, n.1=n1, n.2=n2,
conf.level=1-α),
```

where `smd` is the standardized mean difference argument and d the observed standardized mean difference, `n.1` and `n.2` are the sample size arguments for Groups 1 and 2, with n_1 and n_2 being the sample sizes of Group 1 and Group 2, respectively, and `conf.level` is the confidence interval argument, with $1-\alpha$ being the desired level of confidence. For example, suppose a researcher performs an experiment in which the observed standardized mean difference between the treatment group and the control group is .525, where there were 64 participants in each of the two groups. The `ci.smd()` function could be used as

```
R> ci.smd(smd=.525, n.1=64, n.2=64,
conf.level=.95),
```

which yields a 95% confidence interval,

$$CI_{.95} = [.17152 \leq \delta \leq .87647],$$

where $CI_{.95}$ represents a 95% confidence interval for δ, the population standardized mean difference.[2]

Sample Size Planning

Although there are some functions within R and within certain R packages for planning sample size from a power analytic perspective, where the goal is to obtain results that reach statistical significance, the MBESS package contains functions for planning sample size from the power analytic perspective (e.g., Cohen, 1988; Kraemer & Thiemann, 1987; Lipsey, 1990; Murphy & Myors, 2004) as well as from the accuracy in parameter estimation (AIPE) perspective (e.g., Hahn & Meeker, 1991; Kelley & Maxwell, 2003, 2008; Kelley, Maxwell, & Rausch, 2003; Kelley & Rausch, 2006; Kupper & Hafner, 1989). Whereas the goal of the power analytic approach is to plan sample size so that there is some desired probability of rejecting a false null hypothesis, the goal of the AIPE approach is to obtain a sufficiently narrow confidence interval with some desired probability.

UN EJEMPLO PRACTICO

Tratamiento: Ciproheptadina 3-11 dias

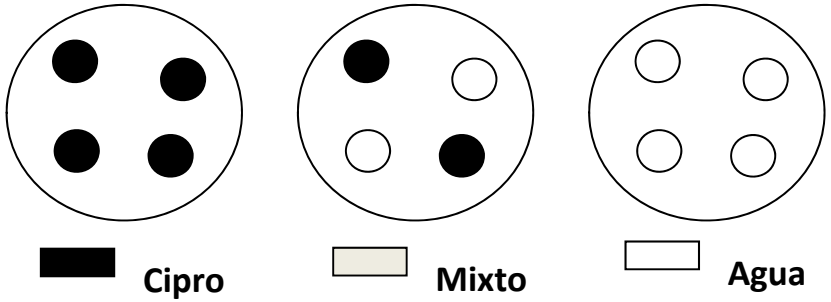
Respuesta inmune celular (PHA) A los 13 días

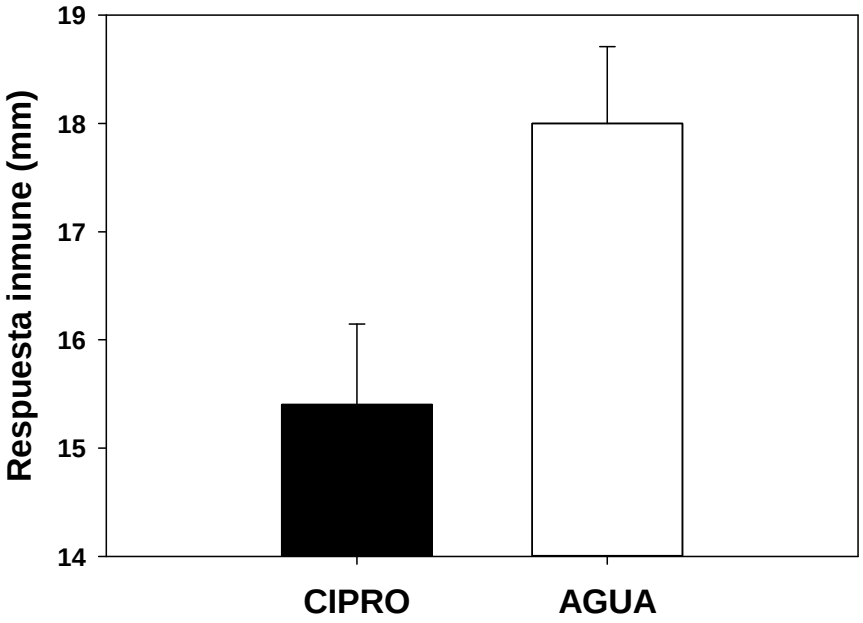
¿Cuántos pollitos de gallina necesitaría para estar razonablemente seguro de que ese efecto no lo produce la ciproheptadina?

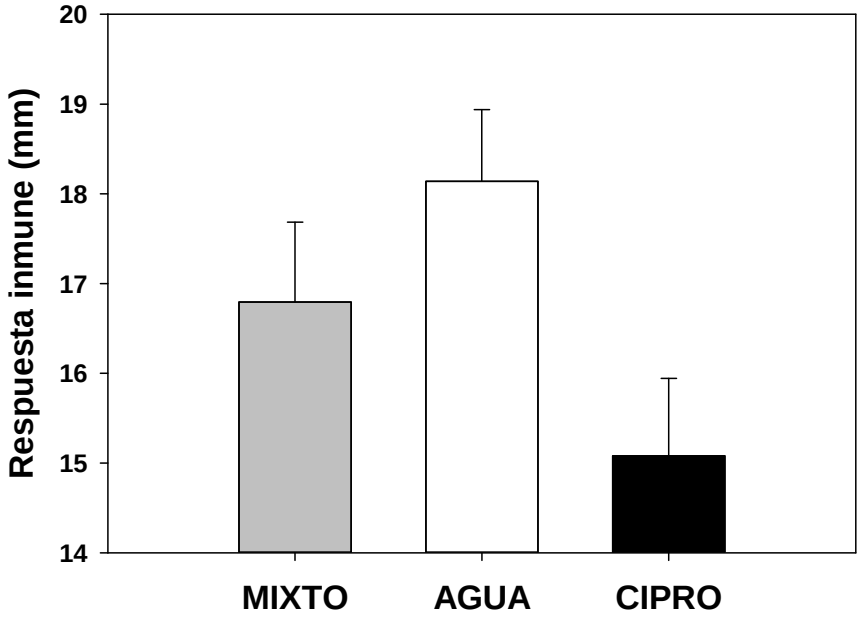
HIPOTESIS: Si la ciproheptadina es la causa, debería detectarse un efecto similar en pollitos de gallina

	media	SD	CV
Control	2.71	1.16	0.43
Experimental	2.34	1.14	0.49

La estructura de factores fijos (fecha de eclosión, tratamiento del nido) y aleatorios (nido, zona) del modelo original no es reproducible en el nuevo experimento







Hacemos una estima a lo bruto del tamaño de efecto sin ningún tipo de estructura

	media	SD	CV
Control	2.71	1.16	0.43
Experimental	2.34	1.14	0.49

https://www.psychometrica.de/effect_size.html

	Group 1	Group 2
Mean	2.71	2.34
Standard Deviation	1.16	1.14
Sample Size (N)	130	125
Effect Size d_{Cohen} g_{Hedges} *	-0.322	

Confidence Coefficient	95%
Confidence Interval	-0.569 - -0.075

https://www.psychometrica.de/effect_size.html

	Group 1	Group 2
Mean	2.71	2.34
Standard Deviation	1.16	1.14
Sample Size (N)	130	125
Effect Size d_{Cohen} g_{Hedges} *	-0.322	

Confidence Coefficient	95%
Confidence Interval	-0.569 - -0.075

https://www.p

Sta

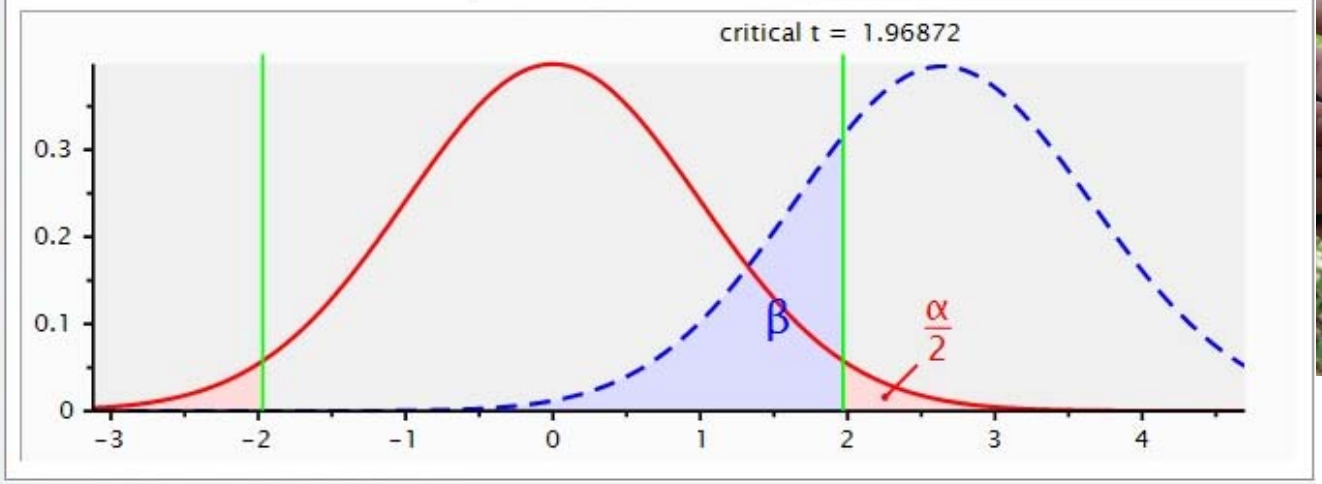
Effect Size

Confidenc

Confiden

File Edit View Tests Calculator Help

Central and noncentral distributions | Protocol of power analyses



Test family: t tests

Statistical test: Means: Difference between two independent means (two groups)

Type of power analysis: A priori: Compute required sample size – given α, power, and effect size

Input Parameters

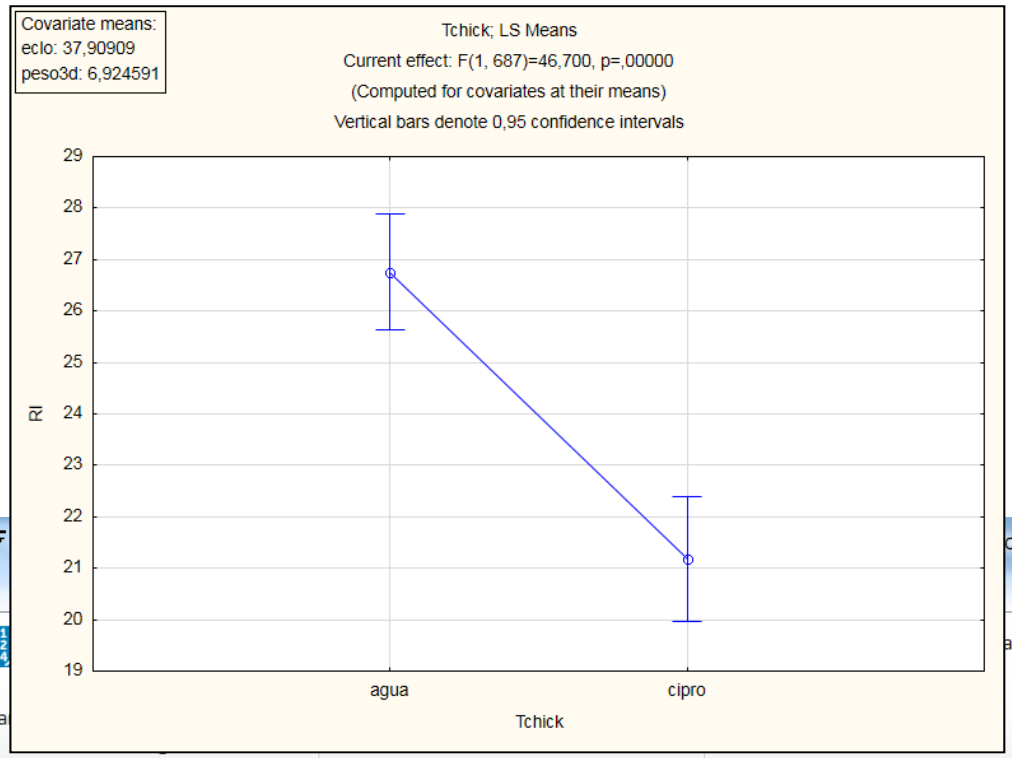
Tail(s)	Two
Determine => Effect size d	0.32
α err prob	0.05
Power (1−β err prob)	0.75
Allocation ratio N2/N1	1

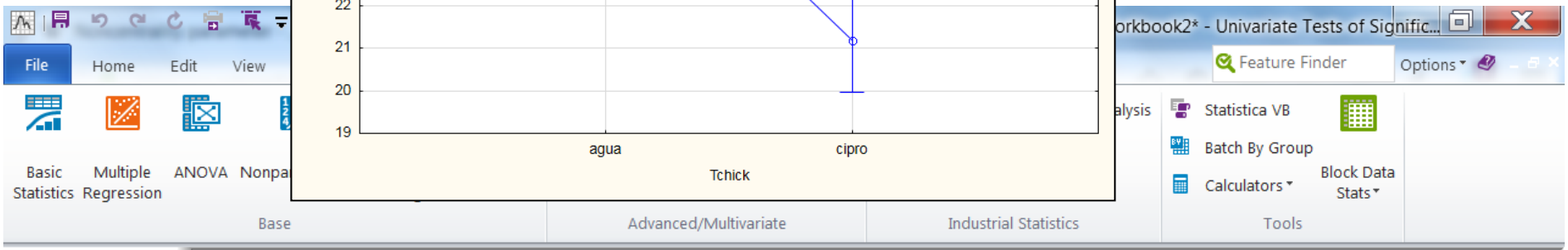
Output Parameters

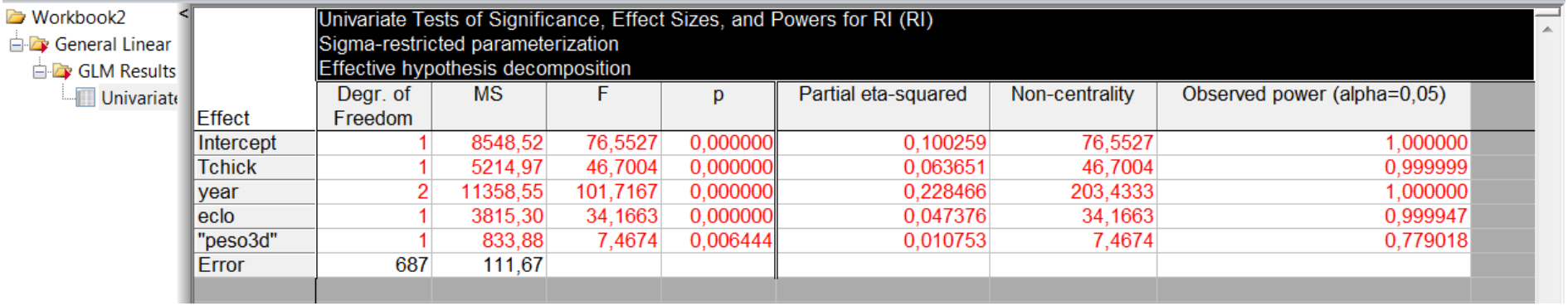
Noncentrality parameter δ	2.6484713
Critical t	1.9687238
Df	272
Sample size group 1	137
Sample size group 2	137
Total sample size	274
Actual power	0.7514793

Para un tamaño de efecto de 0.32 necesitaríamos 137 individuos por tratamiento

Yo espero tener un efecto estandarizado superior a 0.32 porque parte de la varianza es explicable por otros efectos. Un Modelo Lineal me da un eta cuadrado parcial de 0.06



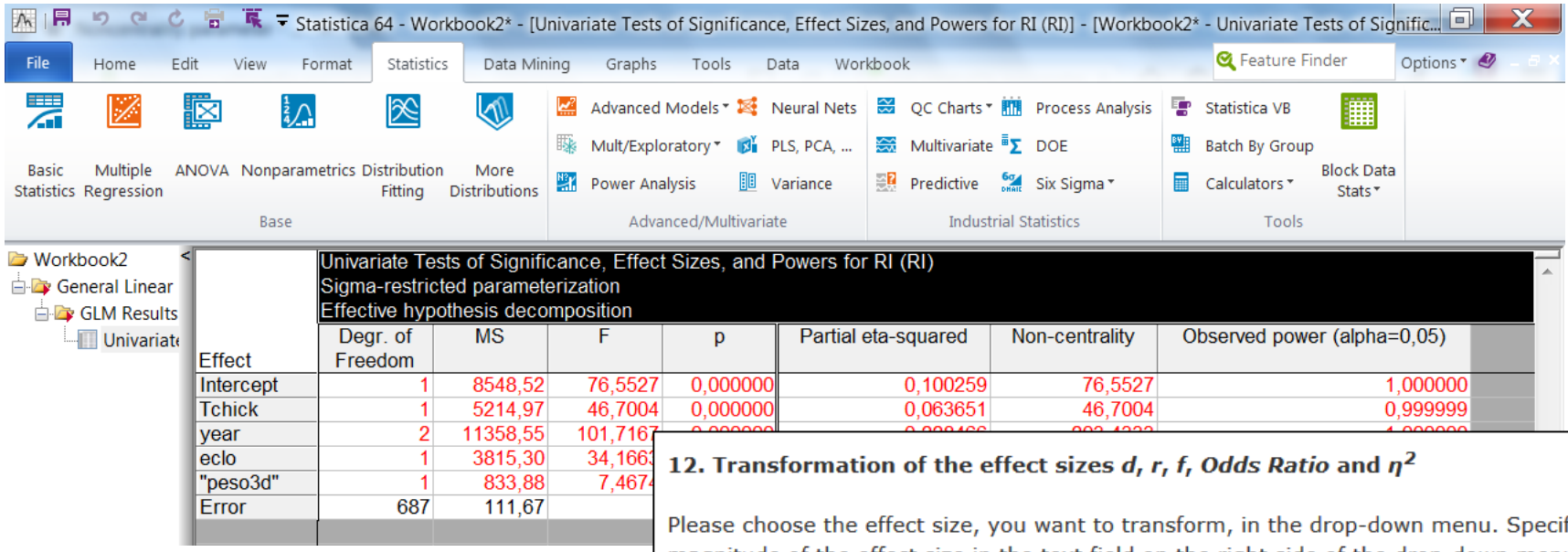




Univariate Tests of Significance, Effect Sizes, and Powers for RI (RI)
Sigma-restricted parameterization
Effective hypothesis decomposition

Effect	Degr. of Freedom	MS	F	p	Partial eta-squared	Non-centrality	Observed power (alpha=0,05)
Intercept	1	8548,52	76,5527	0,000000	0,100259	76,5527	1,000000
Tchick	1	5214,97	46,7004	0,000000	0,063651	46,7004	0,999999
year	2	11358,55	101,7167	0,000000	0,228466	203,4333	1,000000
eclo	1	3815,30	34,1663	0,000000	0,047376	34,1663	0,999947
"peso3d"	1	833,88	7,4674	0,006444	0,010753	7,4674	0,779018
Error	687	111,67					

Yo espero tener un efecto estandarizado superior a 0.32 (podría ser hasta 0.5)



El efecto d equivalente para el modelo lineal sería de 0.5

12. Transformation of the effect sizes *d*, *r*, *f*, *Odds Ratio* and η^2

Please choose the effect size, you want to transform, in the drop-down menu. Specify the magnitude of the effect size in the text field on the right side of the drop-down menu afterwards. The transformation is done according to Cohen (1988), Rosenthal (1994, S. 239) and Borenstein, Hedges, Higgins, und Rothstein (2009; transformation of d in Odds Ratios).

Effect Size	Eta Quadrat	0.06
d		0.5053
r		0.2449
η^2		0.06
f		0.2526
Odds Ratio		2.5005
Number Needed to Treat (NNT)		3.5826

https://www.psychometrica.de/effect_size.html

Yo espero (y además procuro) tener un efecto estandarizado superior a 0.32

12. Transformation of the effect sizes *d*, *r*, *f*, *Odds Ratio* and η^2

Please choose the effect size, you want to transform, in the drop-down menu. Specify the magnitude of the effect size in the text field on the right side of the drop-down menu afterwards. The transformation is done according to Cohen (1988), Rosenthal (1994, S. 239) and Borenstein, Hedges, Higgins, und Rothstein (2009; transformation of d in Odds Ratios).

Effect Size	Eta Quadrat	0.06
d	0.5053	
r	0.2449	
η^2	0.06	
f	0.2526	
Odds Ratio	2.5005	
Number Needed to Treat (NNT)	3.5826	

Menor CV (0.26)
Diseño por bloques
Refinamiento
Experimento en laboratorio

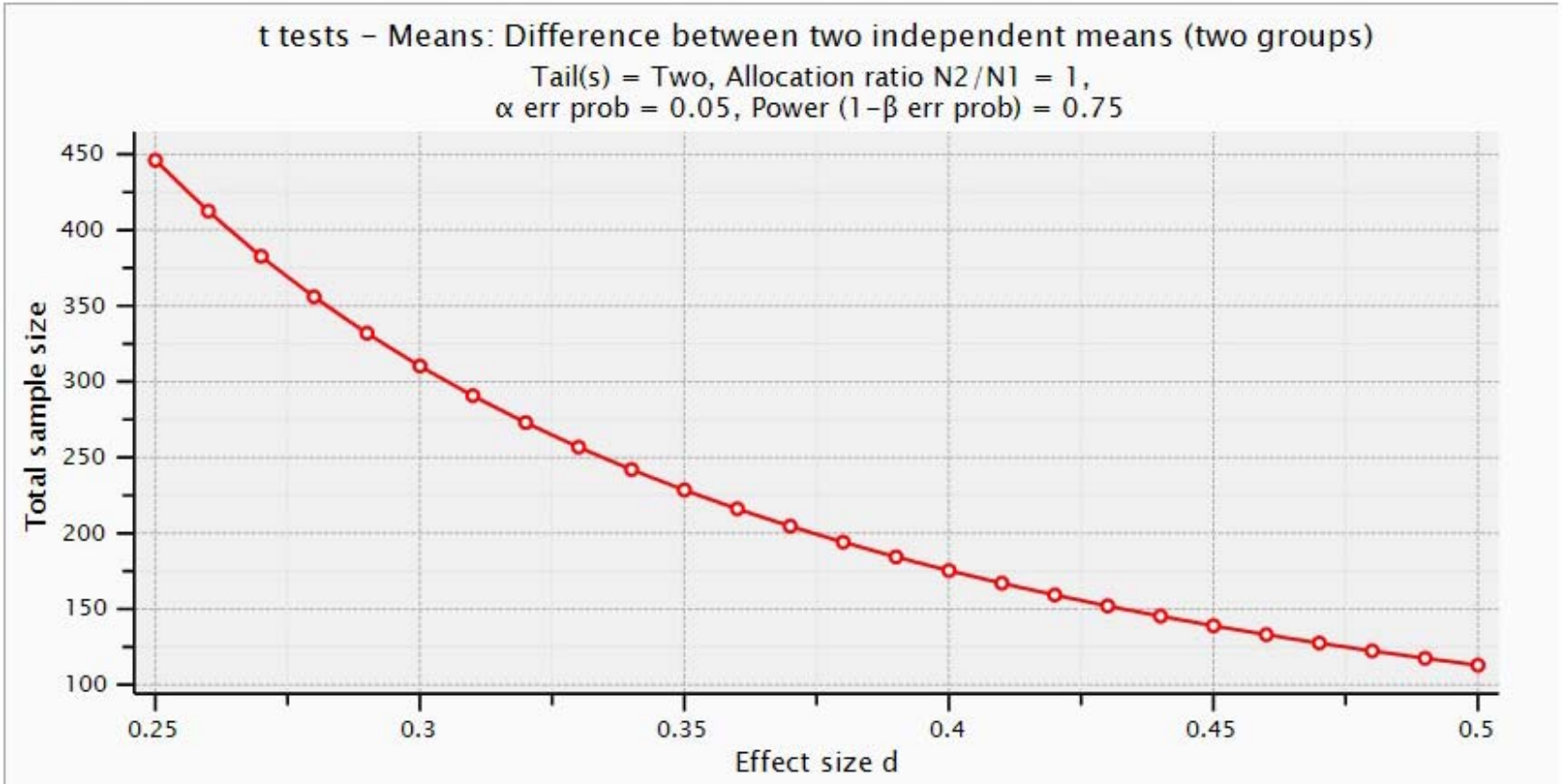
Growth Performance, Carcass Yield, and Immune Competence of Broiler Chickens as Influenced by Dietary Supplemental Zinc Sources and Levels

K. Yogesh, Chandra Deo,
H. P. Shrivastava, A. B. Mandal,
Ashutosh Wadhwa & Indira Singh

GPower - Plot

File Edit View

Graph Table



Plot Parameters

Plot (on y axis) Total sample size ☑ with markers ☐ and displaying the values in the plot

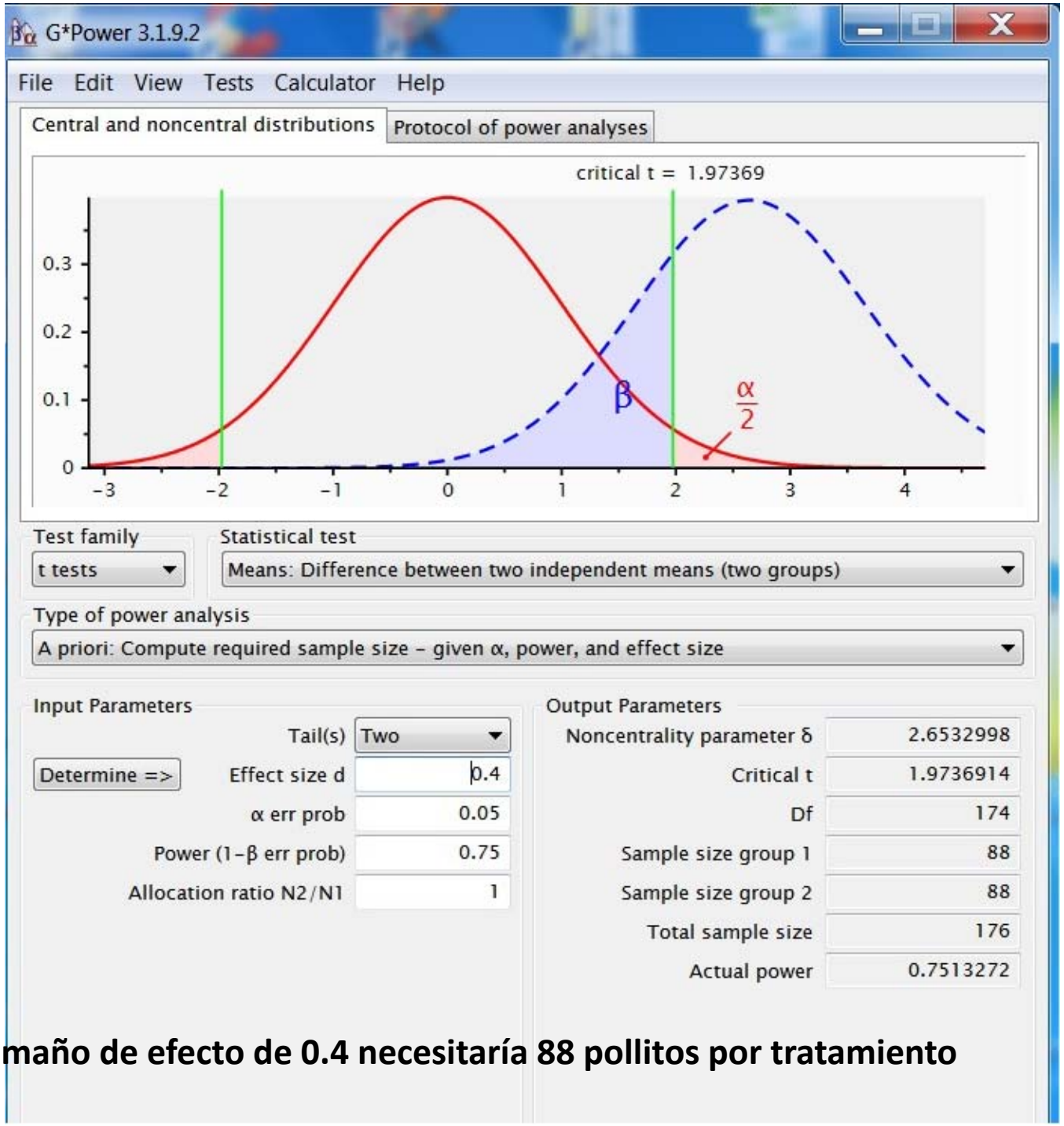
as a function of Effect size d from 0.25 in steps of 0.01 through to 0.5

Plot 1 graph(s) interpolating points

with Power (1−β err prob) at 0.75

and α err prob at 0.05

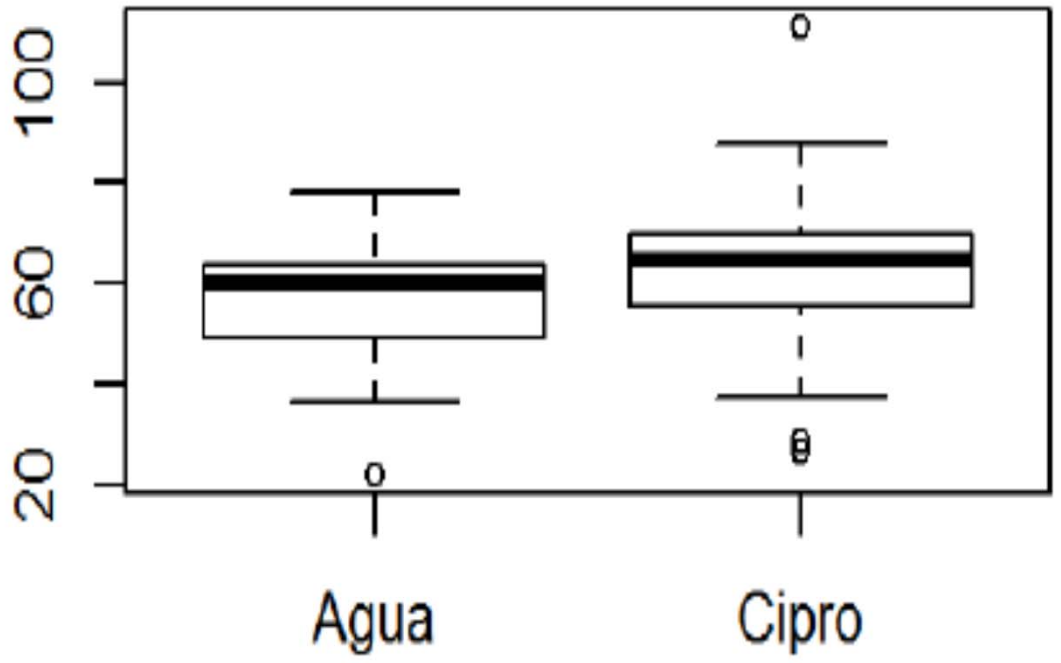
Draw plot



Para un tamaño de efecto de 0.4 necesitaría 88 pollitos por tratamiento

Obtengo un efecto OPUESTO de tamaño 0.46 (0.01 – 0.92 95 % IC) con un primer lote de 40 pollos por tratamiento

```
## Data: d1
## Models:
## model1.null: YR ~ (1 | id1)
## fitF: YR ~ Xb1 + (1 | id1)
##              Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model1.null   3 633.92 640.99 -313.96   627.92
## fitF          4 631.98 641.40 -311.99   623.98 3.9439      1    0.04704 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Effect size Nakagawa & Cuthill (2007): 0.465435 **95 % CI:** 0.0096549 0.9212151

Puedo calcular la potencia “observada” y simular el número de pollos por casilla con SIMR

Methods in Ecology and Evolution

British Ecological Society

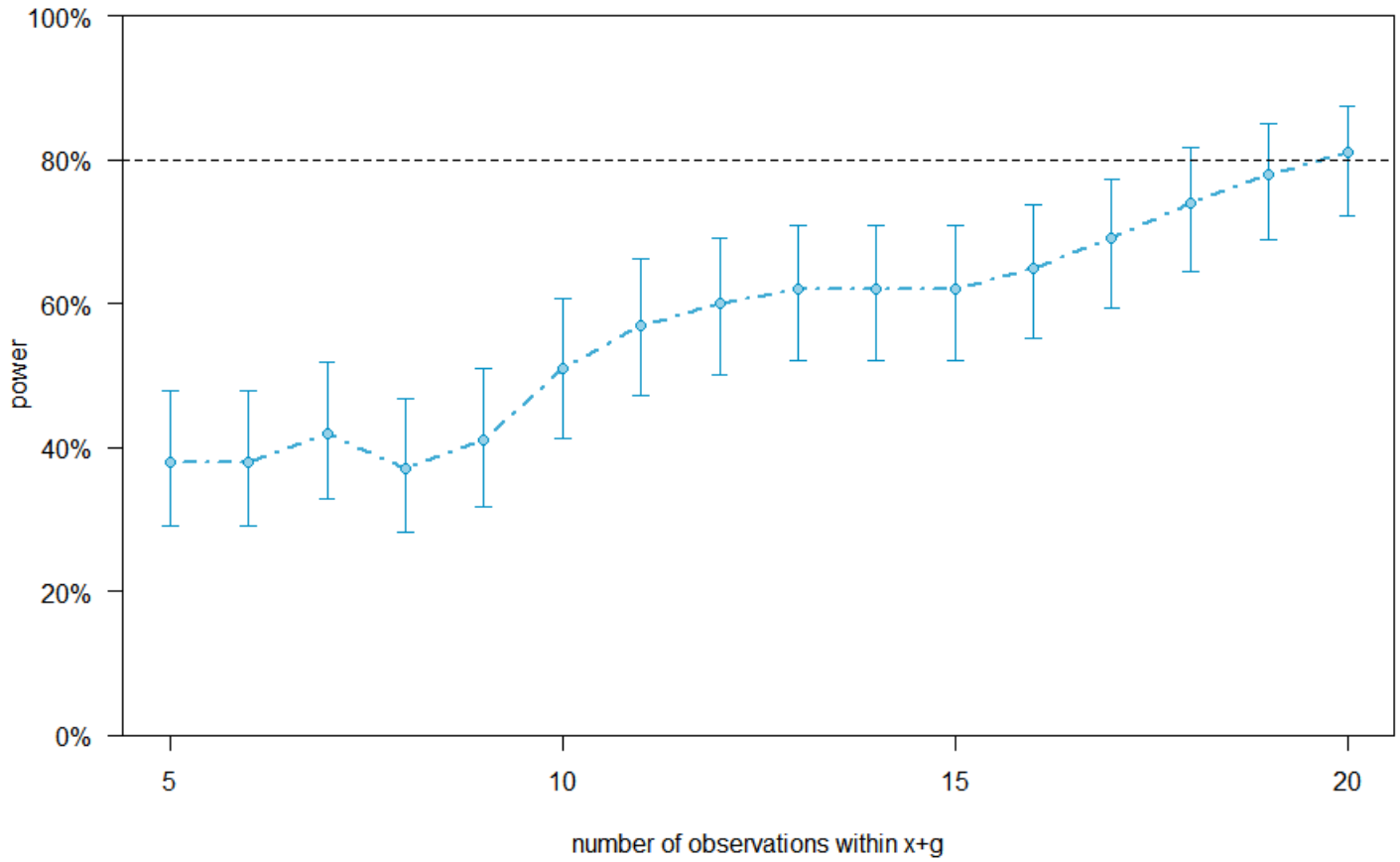
Methods in Ecology and Evolution 2016 doi: 10.1111/2041-210X.12504

APPLICATION

SIMR: an R package for power analysis of generalized linear mixed models by simulation

Peter Green* and Catriona J. MacLeod

Landcare Research, Private Bag 1930, Dunedin 9054, New Zealand



O simular el numero de jaulas manteniendo constante el numero de pollos por jaula

Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2016

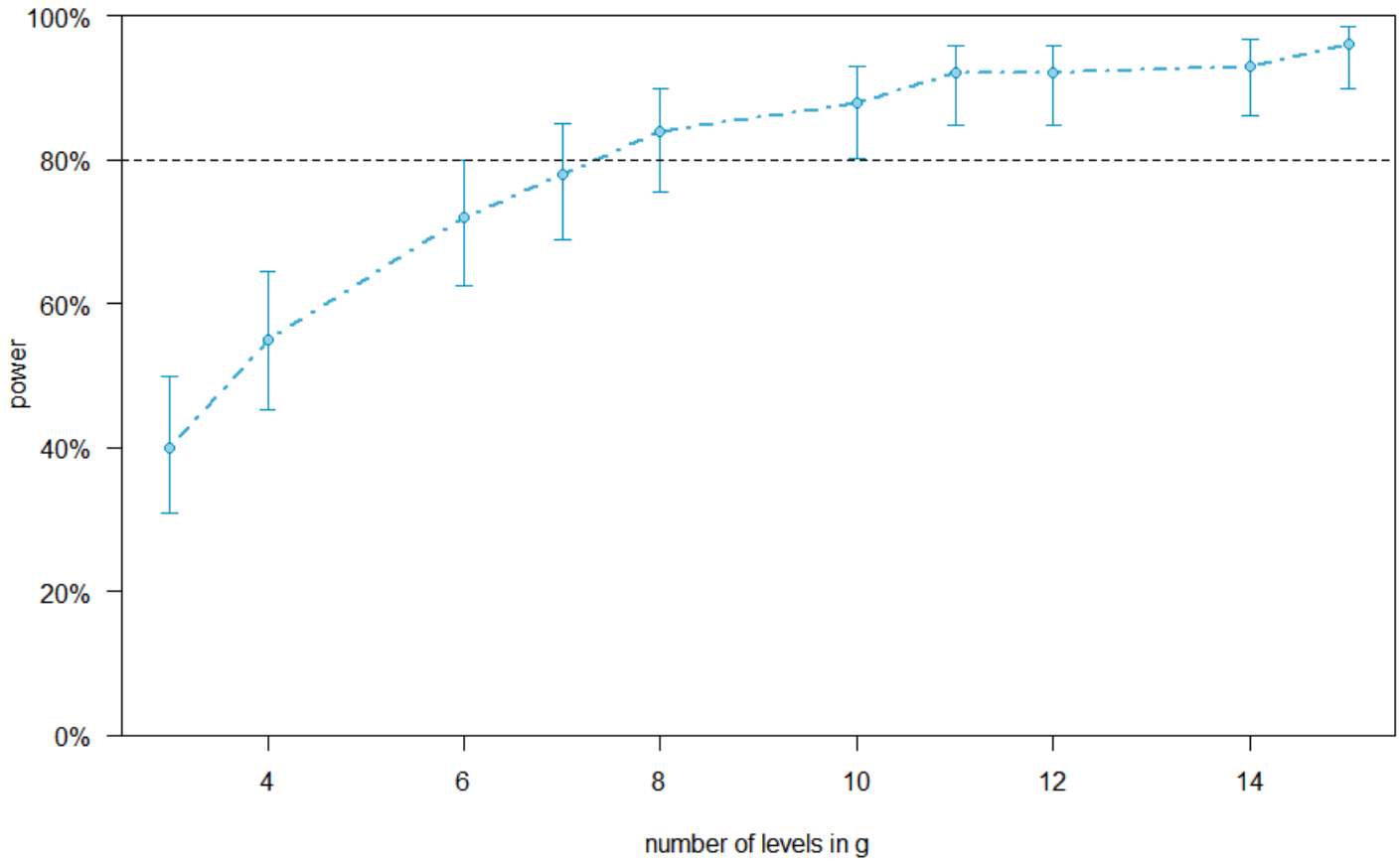
doi: 10.1111/2041-210X.12504

APPLICATION

SIMR: an R package for power analysis of generalized linear mixed models by simulation

Peter Green* and Catriona J. MacLeod

Landcare Research, Private Bag 1930, Dunedin 9054, New Zealand



Puedo calcular mi tamaño de efecto en gallinas igual que la estima que hice para diseñar el estudio https://www.psychometrica.de/effect_size.html

	media	SD	N
Control	8.98	1.91	37
Experimental	10.03	2.81	39

	Group 1	Group 2
Mean	89.85	100.31
Standard Deviation	19.06	27.89
Sample Size (N)	37	39
Effect Size d_{Cohen} g_{Hedges} *	0.436	

Confidence Coefficient	95%
Confidence Interval	-0.019 - 0.891

Effect size Nakagawa & Cuthill (2007): 0.465435 **95 % CI:** 0.0096549 0.9212151

O calcular los IC y una estima AIPE del tamaño de muestra para una exactitud de 0.35 usando el paquete MBESS

R> ci.smd(smd=0.465, n.1=37, n.2=39, conf.level=1-.05)

$Lower.Conf.Limit.smd
[1] 0.007494684
$smd
[1] 0.465 $
Upper.Conf.Limit.smd
[1] 0.9194494

R> ss.aipe.smd(delta=.465, conf.level=.95, width=.45, degree.of.certainty= 0.95)

[1] 160

R> ss.aipe.smd(delta=.465, conf.level=.95, width=.35, degree.of.certainty= 0.95)

[1] 263

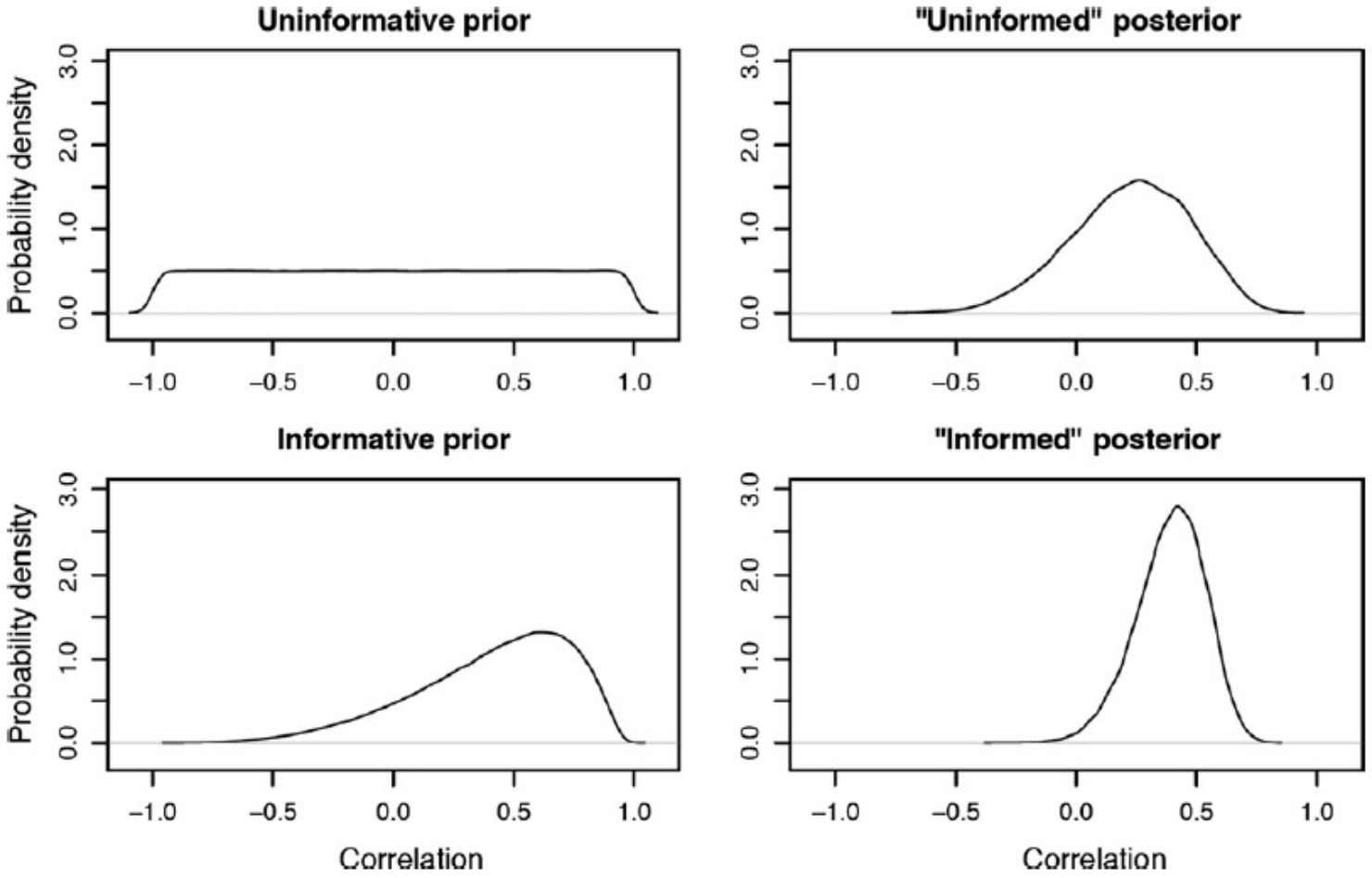
Package 'MBESS'

September 23, 2016

Type Package
Title The MBESS R Package
Version 4.1.0
Date 2016-09-21
Maintainer Ken Kelley <kkelley@nd.edu>
Depends R (>= 3.2.0), stats

Análisis Bayesiano

$$\Pr(\theta|D) = \frac{\Pr(\theta) \times \Pr(D|\theta)}{\Pr(D)},$$



Treatment 2 is more cost-effective than Treatment 1 if $\beta(K) > 0$

Analysis Objective

- Frequentist formulation
  - We wish to reject the null hypothesis that $\beta(K) = 0$, at the $100(1 - \omega)\%$ level of significance
  - e.g. $\omega = 0.95$ corresponds to usual 5% test
- Bayesian formulation
  - We wish to have at least a $100\omega\%$ posterior probability that $\beta(K) > 0$

Treatment 2 is more cost-effective than Treatment 1 if $\beta(K) > 0$

Design Objective

- Frequentist formulation
  - We want sample sizes large enough to give $100\delta\%$ power to reject the null hypothesis that $\beta(K) = 0$, when the true value of $\beta(K)$ has some assumed alternative value
- Bayesian formulation
  - We want sample sizes large enough to give a $100\delta\%$ prior probability of achieving the desired posterior probability that $\beta(K) > 0$

Treatment 2 is more cost-effective than Treatment 1 if $\beta(K) > 0$

- Putting the two frequentist objectives together:

  "We wish to have 70%power to reject the null hypothesis that $\beta(K)=0$ *at the 5% level, if*

- Putting the two Bayesian objectives together:

  "We wish to be 70% sure of obtaining a 95% posterior probability that $\beta(K) > 0$*"*