



Simpósio AMP



23 Novembro 2013

14:30-16:00

Os novos artigos científicos

Moderação - António Vaz Carneiro, João Massano

14:30-14:50

• **A nova bibliometria, as novas métricas da moderna publicação científica - Helena Donato**

14:50-15:10

• **Medical Writing. Tipos de estudo científicos e nível de evidência - Francisco Batel Marques**

15:10-15:30

• **A Bioestatística salva vidas? - Vitor Rodrigues**

15:30-15:50

• **As palavras chave - Fernando Fernandez-Llimós**

15:50-16:00

Discussão

Vitor Rodrigues
Faculdade de Medicina – Universidade de Coimbra
vrodrigues@fmed.uc.pt

A Bioestatística salva vidas?

A Bioestatística salva vidas?

(Provavelmente)

Não

A Bioestatística salva vidas?

Mas ajuda!!!!!!

A Bioestatística salva vidas?

Sobretudo SE “bem feita”

Bioestadística!!!!!!!



Bioestadística!!!



Bioestadística 😊



Bioestatística de sucesso = Culinária de sucesso

- Bons ingredientes (em quantidade e em qualidade)
- Desenho culinário (Pantagruel) adequado
- Para um objectivo claro
 - Para a sogra?
 - Para o/s namorado/a?
 - Para os/as amigos/as?
 - Para o/a chefe?
 - Para uso próprio?
- Ervas aromáticas/condimentos para melhorar e/ou adequar o sabor e não para disfarçar a má qualidade dos alimentos e o erro na confecção

O mau uso da bioestatística pode:

- É um problema potencialmente grave
- Pode levar a:
 - Resultados distorcidos
 - Conclusões incorrectas
 - Desperdício de recursos valiosos
- Não ético
- Pode provocar graves consequências clínicas

When the researcher says:

The demographic characteristics of the nonresponders were similar to those of the rest of the sample

He or she really means:

All we really had on them were age and sex

Exemplos de erros estatísticos em artigos publicados:

Error	Error rate N (proportion)
Power analysis not provided	55/55 (1.0)
Incorrect use of statistical test for comparing three or more groups for differences	21/28 (0.75)
Incorrect presentation of P value	36/54 (0.66)
Incorrect choice of the statistical test	34/55 (0.62)
Incorrect interpretation of correlation analysis	11/20 (0.55)
Incorrect use or presentation of descriptive analysis	19/55 (0.35)
Incorrect interpretation of P value	12/54 (0.22)

Ana-Maria Šimundić*, Nora Nikolac. Statistical errors in manuscripts submitted to Biochemia Medica journal. Biochemia Medica, 2009;19(3):294-300

When the researcher says:

Agreement between the raters was acceptable

He or she really means:

The agreement was so bad that we don't dare to include the actual number in the paper

Mais exemplos:

- McKinney WP, et al. The inexact use of Fisher's Exact Test in six major medical journals. JAMA. 1989;261:3430-3.
- Kanter MH, Taylor JR. Accuracy of statistical methods in Transfusion: a review of articles from July/August 1992 through June 1993. Transfusion. 1994;34:697-701.
- Kuo YH. Extrapolation of correlation between 2 variables in 4 general medical journals. JAMA. 2002 Jun 5;287(21):2815-7.
- Nagele P. Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. Br J Anaesth. 2003 Apr; 90(4):514-6.
- Simundic AM, Nikolac N, Topic E. Methodological issues in genetic association studies of inherited thrombophilia. Clin Appl Thromb Hemost. 2009;15(3):327-33.
- Simundic AM, Nikolac N. Most common statistical errors of articles submitted to Biochem Med. 2009;19(3):294-300.

When the researcher says:

The questionnaire was circulated to a panel of experts to establish face validity

He or she really means:

Our friends liked it...and the bottle of scotch we included

Alguns erros básicos:

- Amostragem
- Representação gráfica
- Apresentação dos dados
- Escolha do teste estatístico
- Valor de p
- Conclusões, causalidade
- Teste de hipóteses múltiplas

When the researcher says:

The rate of lung cancer among the hourly rate employees was significantly higher, which may be caused by excess PCBP exposure.

He or she really means:

It might also be caused by obvious things like smoking and social class, but I'm interested in PCBP today.

Talvez o erro maior:

PERGUNTA (cada vez menos habitual):

Tenho aqui estes dados! O que faço com eles?

RESPOSTA:

..... (visado pela comissão de censura)

When the researcher says:

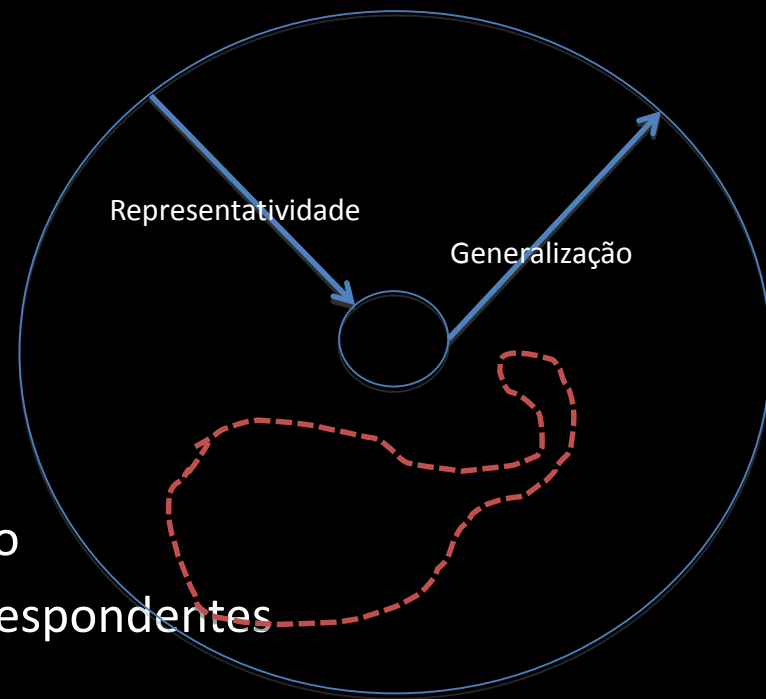
The response rate was 60%, which is acceptable for studies of this type.

He or she really means:

However, the study itself was so bad that even a 100% response rate wouldn't have saved it.

Amostragem:

- Amostras de conveniência assumidas como aleatórias
- Não calcular o n imediatamente
 - Quantos preciso? “mínimo de 30”
 - Quantos preciso? Os que se quiserem
 - Quantos preciso? Qual o limiar de “incerteza, ...”
- Critérios de exclusão e de inclusão
- Unidades amostrais não definidas
- Desigualdade de sub-grupos
- Pouco cuidado com os parâmetros do cálculo
- Não quantificação (e qualificação) dos não-respondentes
- Falta de atenção ao “abandono” do estudo



When the researcher says:

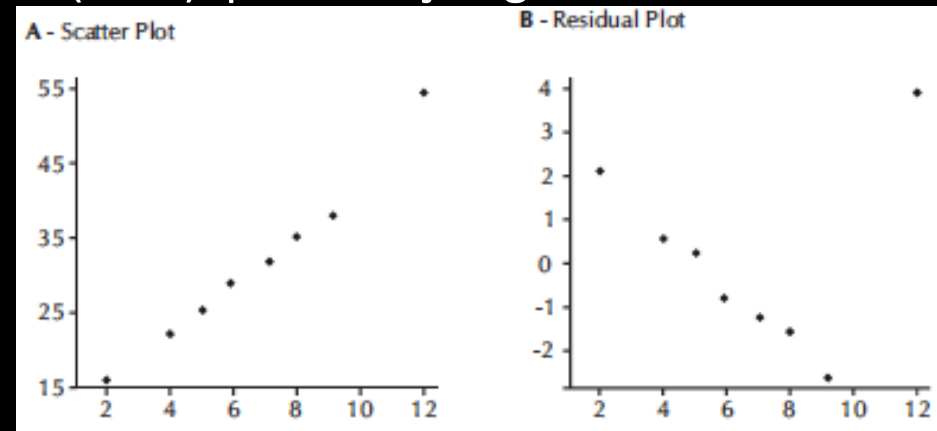
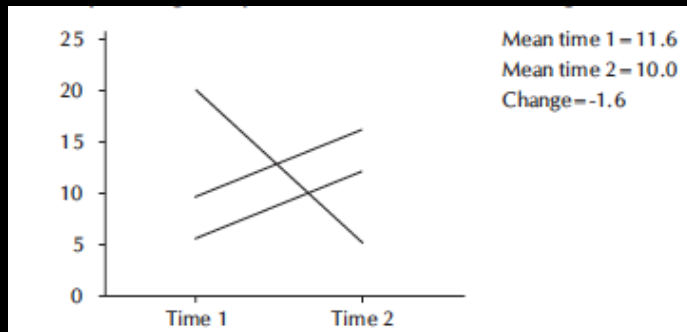
In a case series of 12 patients, 9 showed clinically significant improvement on the experimental drug

He or she really means:

With the help of the drug company representative, I judged which patients got better under my care

Representação gráfica:

- Não iniciar a análise por construir uma (auto)apresentação gráfica dos dados
- Não “olhar com profundidade)
- Uso de gráficos inapropriados



When the researcher says:
The correlation was highly significant ($p < 0.0001$).

He or she really means:
With 10,000 subjects, *any* correlation is highly significant

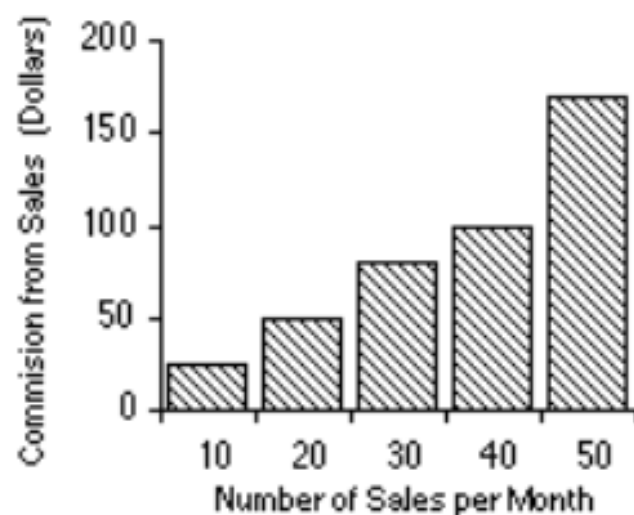
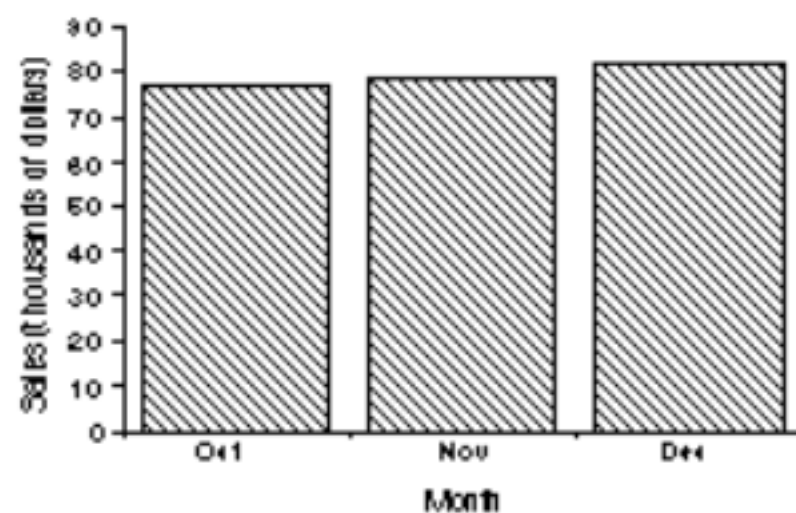
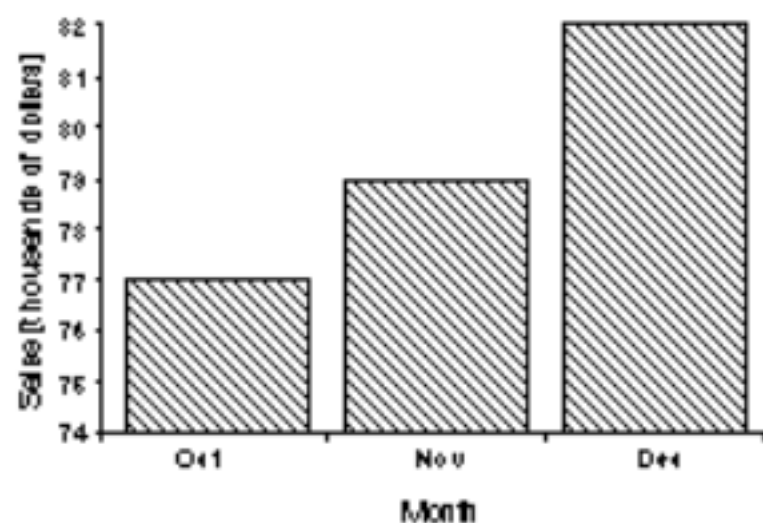


Figure A

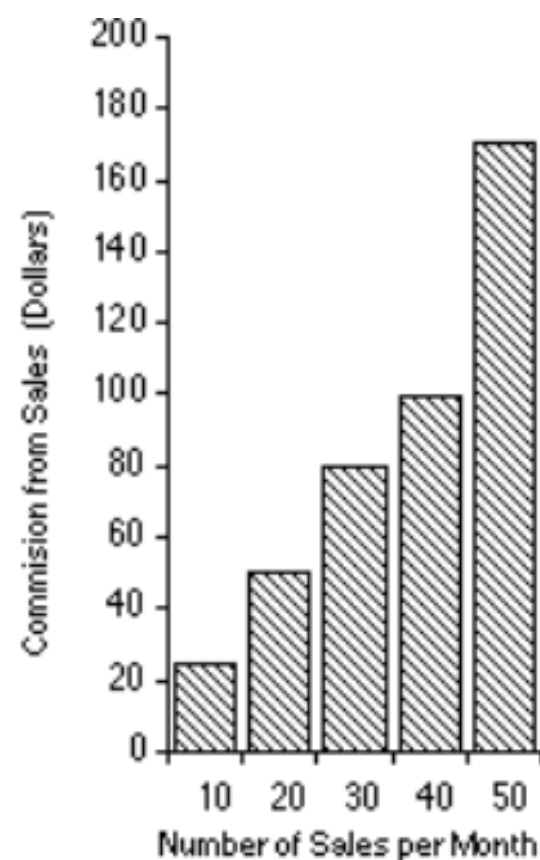
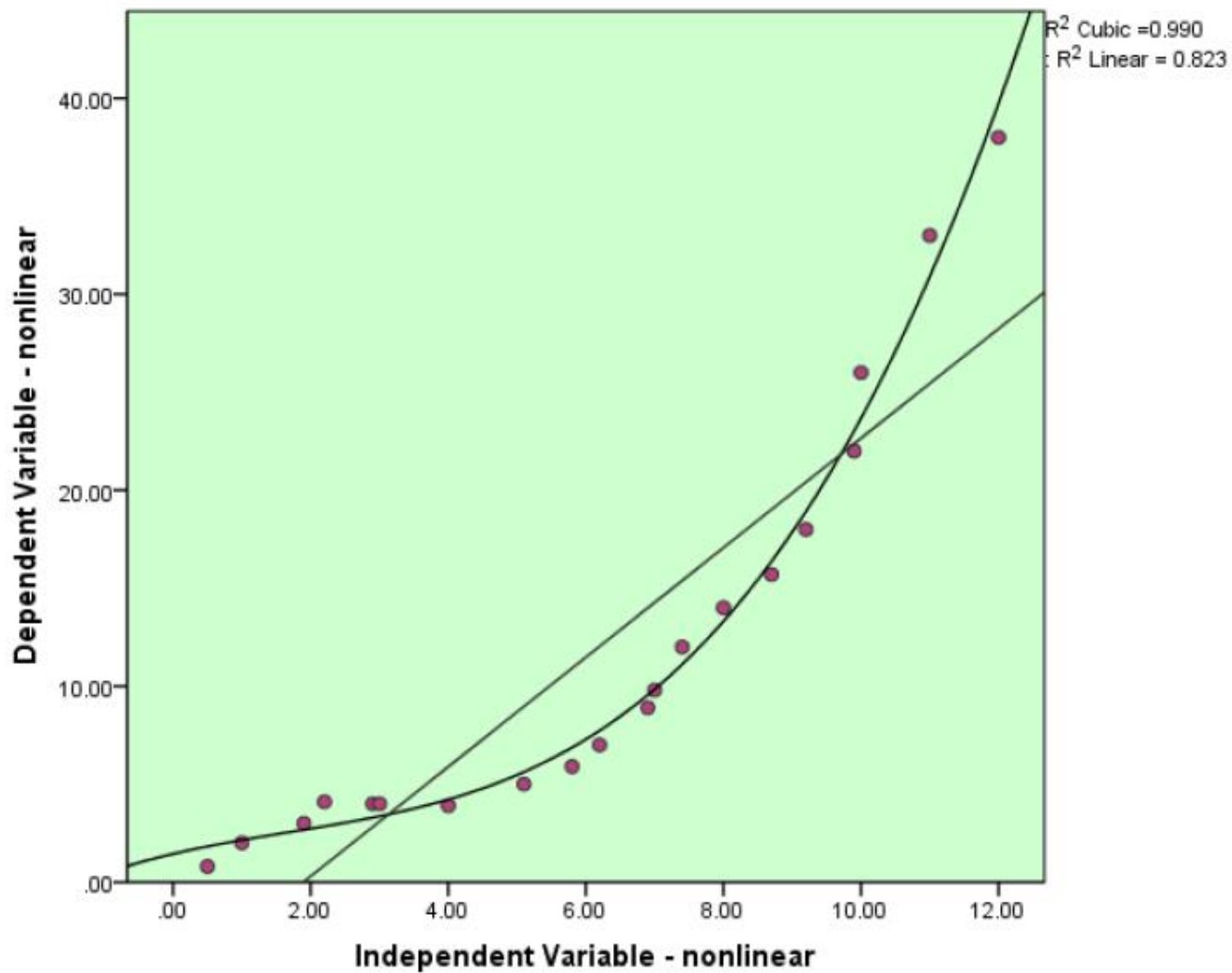


Figure B



When the researcher says:

Further research is required to clarify the results

He or she really means:

I haven't a clue what it all means

Apresentação dos dados:

- Mais comum:
 - Fornecer a média sem apresentar a variabilidade
 - Fornecer SE em vez do SD para descrever os dados
 - Usar média (SD) para descrever dados não-normais
 - Incapacidade de definir a noção de + para descrever a variabilidade ou uso em barras de erro sem label
- Cuidado com as percentagens: se acontecer 1 em 4 – 25%
- Distribuição paramétrica (média + SD)
- Distribuição não-paramétrica (mediana, IQR)

When the researcher says:

A trend was noted

He or she really means:

The statistical test was not significant

Escolha do teste estatístico:

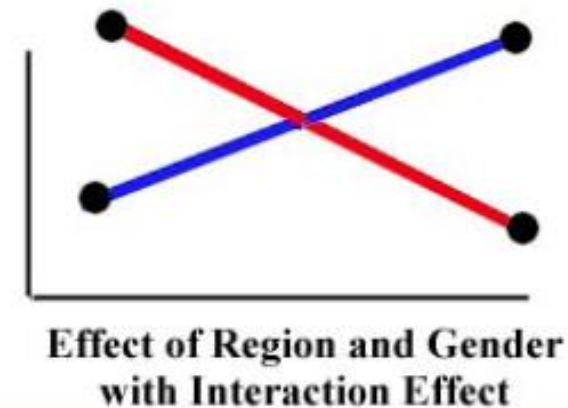
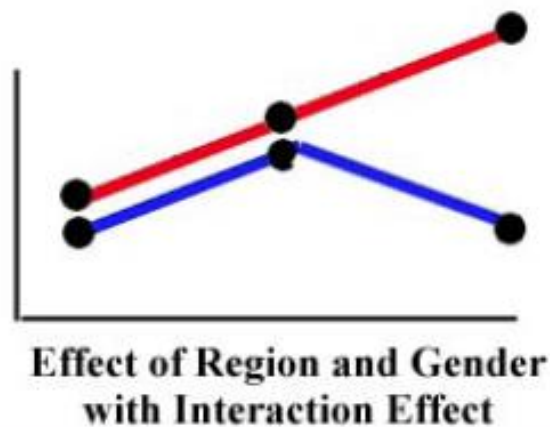
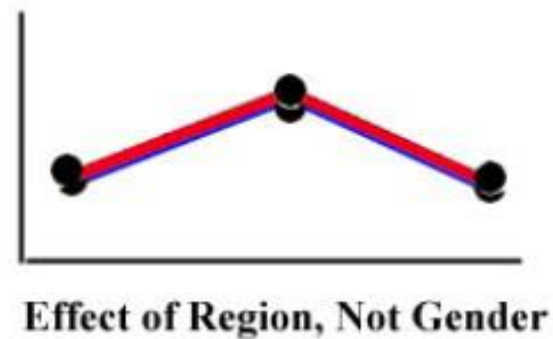
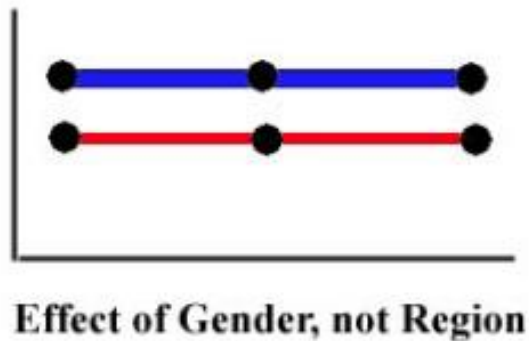
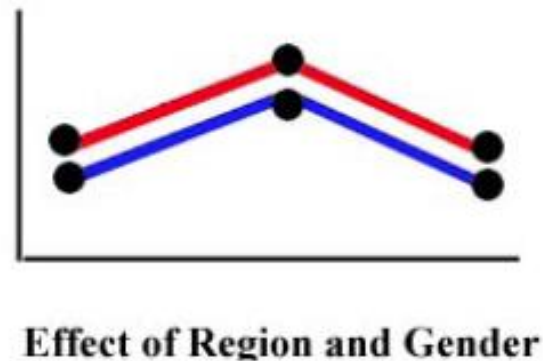
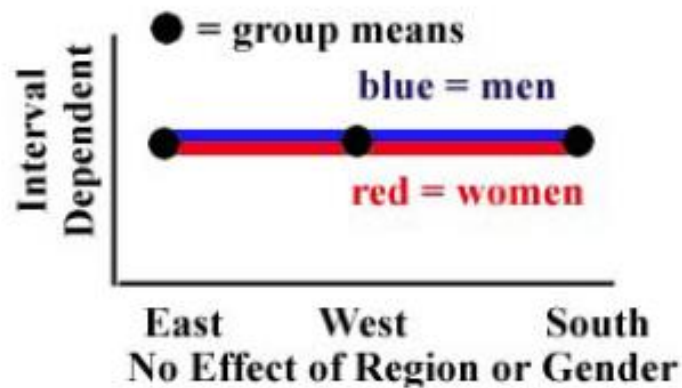
- Uso de testes estatísticos errados
 - Incompatibilidade de teste estatístico com os dados examinados
 - Testes não-emparelhados para dados emparelhados e vice-versa
 - Uso inadequado de métodos paramétricos
 - Uso de testes inadequados para a hipótese em investigação
- Inflação do erro tipo I
 - Incapacidade de incluir uma correção para uma comparação múltipla
 - Análises post-hoc inadequadas
- Erros típicos com teste t-Student
 - Incapacidade para provar suposições de testes
 - Tamanhos amostrais desiguais em testes emparelhados
 - Não identificação de interações (Anova)
- Erros típicos com testes qui-quadrado
 - Não uso de correção de Yates em números pequenos
 - Uso de χ^2 quando as frequências esperadas são <5
 - Confusão entre teste de χ^2 para independência versus homogeneidade
- Não uso de técnicas multivariadas para ajustar para variáveis confundentes

When the researcher says:

The study was a single-blind trial

He or she really means:

Everybody knew who was getting what except the poor patient



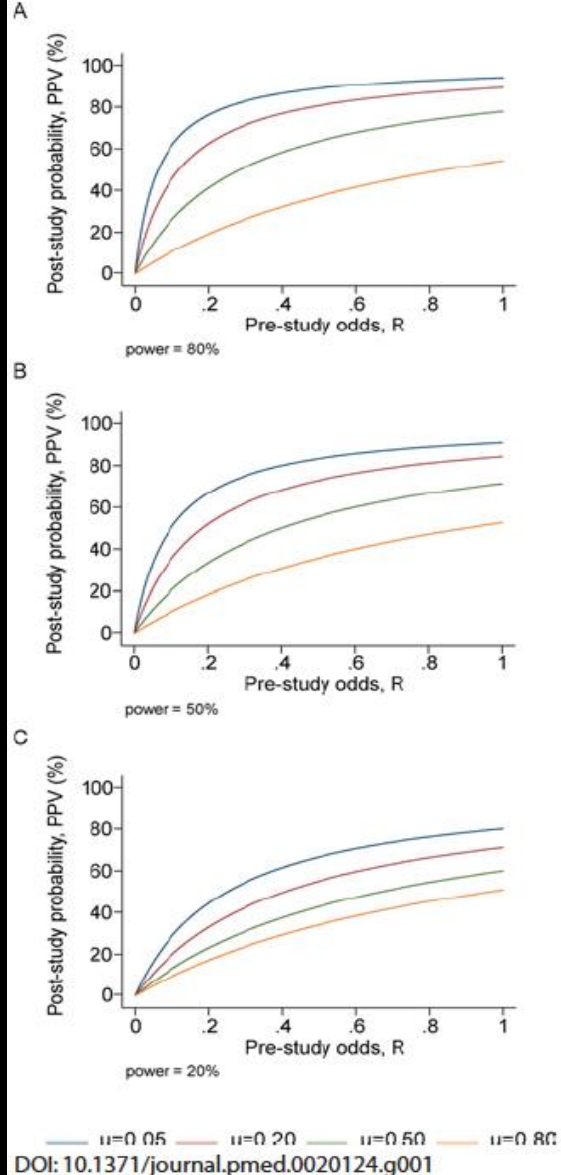


Figure 1. PPV (Probability That a Research Finding Is True) as a Function of the Pre-Study Odds for Various Levels of Bias, u . Panels correspond to power of 0.20, 0.50, and 0.80.

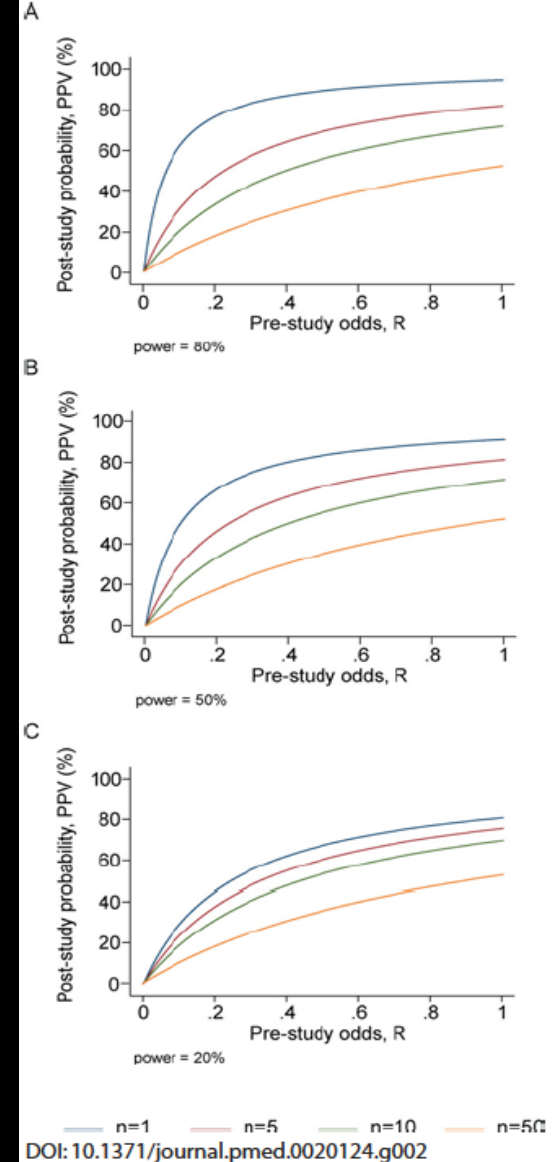


Figure 2. PPV (Probability That a Research Finding Is True) as a Function of the Pre-Study Odds for Various Numbers of Conducted Studies, n . Panels correspond to power of 0.20, 0.50, and 0.80.

$$p \leq 0,05$$



TESTES DE HIPÓTESES

HIPÓTESE NULA **não** há diferenças
não há relação

.....

É DEVIDA AO ACASO

É VERDADE QUE ... ?

É VERDADE -> **CONFIRMAMOS** H_0

É MENTIRA -> **REJEITAMOS** H_0

A QUE NÍVEL DE SIGNIFICÂNCIA ?

TESTES DE HIPÓTESES

HIPÓTESE NULA **não** há diferenças
não há relação

.....

É DEVIDA AO ACASO

É VERDADE QUE ... ?

É VERDADE -> **CONFIRMAMOS** H_0

É MENTIRA -> **REJEITAMOS** H_0

QUAL A MARGEM DE ERRO QUE ADMITIMOS ?

ERRO TIPO I

**Erro cometido quando rejeitamos H_0
sendo esta verdadeira**

Probabilidade de cometer um erro tipo I - α

Probabilidade de cometer um erro tipo II - β

Potência de um teste - $1-\beta$

Nível de significância de um teste - α

ERRO TIPO II

**Erro cometido quando aceitamos H_0
sendo esta falsa**

	Ho confirmada	Ho rejeitada
Ho é verdadeira	Decisão correcta	Erro tipo I
Ho é falsa	Erro tipo II	Decisão correcta

p:

- Confusão entre hipótese nula vs hipótese alternativa e conclusões “inversas”
- Apresentação de p (apenas)
 - Apresentar a diferença absoluta entre grupos e 95% IC
 - Apresentar o teste estatístico e os graus de liberdade
- Uma cauda ou duas caudas?
- Confusão entre significância estatística e significância clínica/biológica
- Não identificar o tamanho amostral como causa de (não)significância estatística

When the researcher says:

Data were analyzed using the Schmedlap-Scheisskopf test.

He or she really means:

We tried the usual tests, but they didn't give significant results

p:

- O risco de náuseas e vômitos pós-operatórios foi superior no grupo placebo comparativamente ao grupo de doentes tratados com dexametasona ($p=0,018$):
- O risco de náuseas e vômitos pós-operatórios foi superior no grupo placebo comparativamente ao grupo de doentes tratados com dexametasona:
OR: 4,5, 95% IC: 4,15-5,35, $p=0,018$
- O risco de náuseas e vômitos pós-operatórios foi superior no grupo placebo comparativamente ao grupo de doentes tratados com dexametasona:
OR: 1,01, 95% IC: 1,009-14,821, $p=0,018$

When the researcher says:
A one-tailed test was used

He or she really means:
The results wouldn't be significant with a two-tailed test

Más interpretações comuns de valores de p significativos

ENGANO: Significância estatística (ex. $p \leq 0,05$) quer dizer que os resultados entre os grupos são diferentes (e não devidos a variação casual).

Pensar em :

1. As diferenças são clinicamente relevantes?
 1. Os ensaios grandes conseguem facilmente diferenças estatísticas que não têm consequências práticas
2. Um valor de p pequeno não corrige o erro sistemático (viés)
 1. Desenhos com desenho pobre podem demonstrar significância estatística, mas conduzem a inferências errôneas
3. Valores de p pequenos apenas significam que as diferenças são, com menor probabilidade, devidas a variação aleatória (acaso)
 1. Um valor de $p \leq 0,001$ ou que $0,0001$, que indica uma muito menor probabilidade de variação aleatória, não é importante
 2. A demarcação de que um valor de $p < 0,05$ é significativo é uma convenção, mas que é arbitrária

Más interpretações comuns de valores de p significativos

ENGANO: Falta de significância estatística significa que os resultados não são importantes.

Pensar em :

1. Os intervalos de confiança podem incluir sobretudo valores que são importantes nos cuidados ao doente (e aproximar significância estatística); para intervenções de baixo risco isto pode ser de evidência suficiente?
2. O estudo é ele próprio um outlier?
 1. Considerar outros estudos/dados que estejam disponíveis
3. Avaliar o desenho do estudo
 1. Existem características amostrais que a fazem diferente da população? Existem outras causas de viés?
 2. Existe uma discussão apropriada sobre o poder do estudo?

Testes de hipóteses múltiplas:

- Se testarmos hipóteses múltiplas, poderemos eventualmente detectar diferenças significativas
- Será real? Será puro acaso?
- Na realidade, se efectuarmos 20 testes no mesmo conjunto de dados, é expectável que ocorra pelo menos 1 erro Tipo 1 (alpha) (se $\alpha=0,05$)

When the researcher says:

A retrospective study was conducted

He or she really means:

We had all these data sitting around and needed some fast publications

Este problema ocorre quando:

- Se testa equivalência de grupos nas características baseline
- Se realizam múltiplas comparações pair-wise
- Se testam múltiplos endpoints
- Se realizam análises secundárias ou em subgrupos
- Se realizam análises interinas em dados acumulados (um endpoint em diversos momentos t_n)

Tom Lang, Twenty statistical errors even YOU can find in biomedical research articles
CMJ, 2004;45(4):361-370

When the researcher says:

One possible explanation for these results is . . .

He or she really means:

I can only think of one.

Distribuição Normal (Gauss)

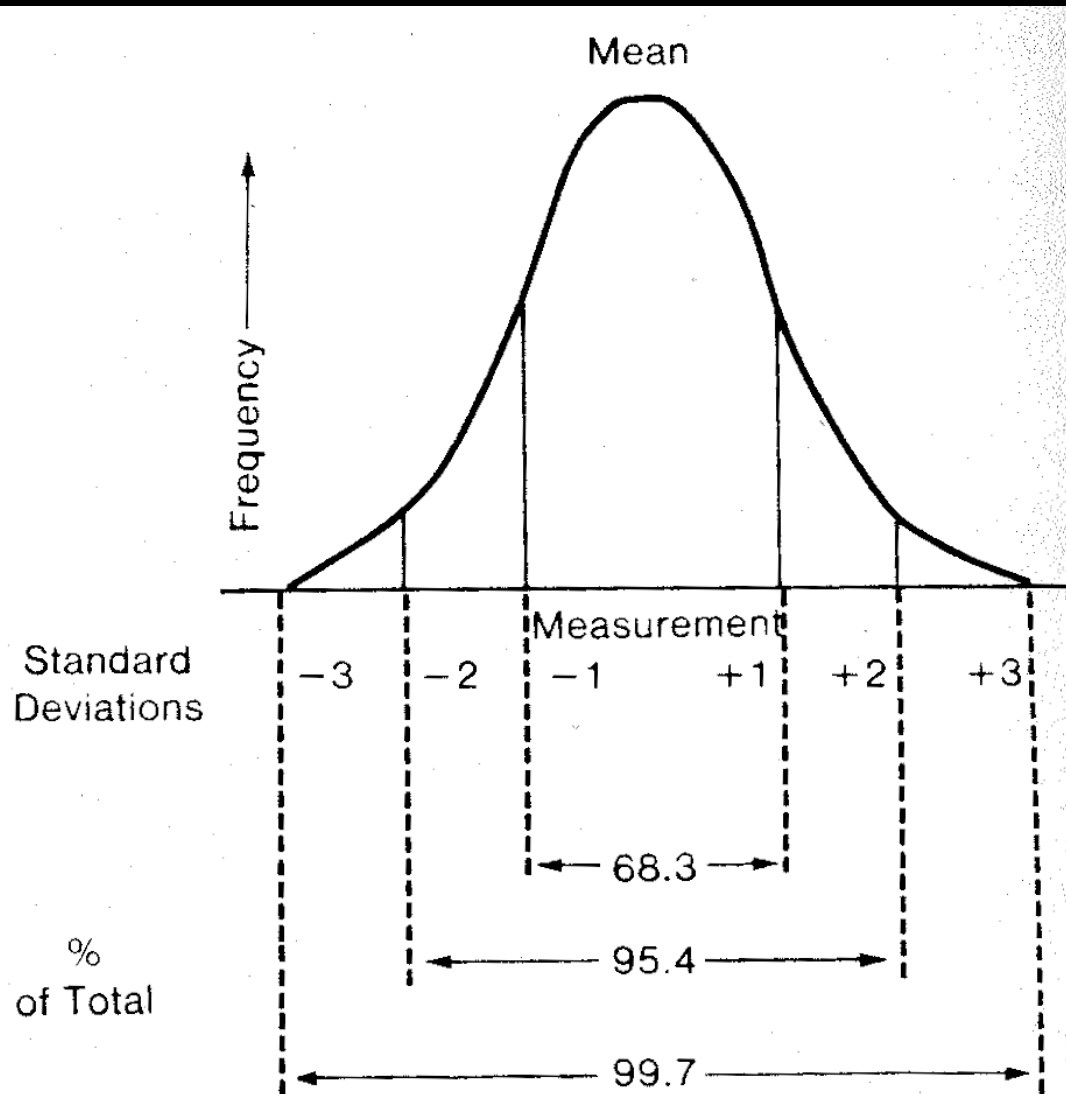
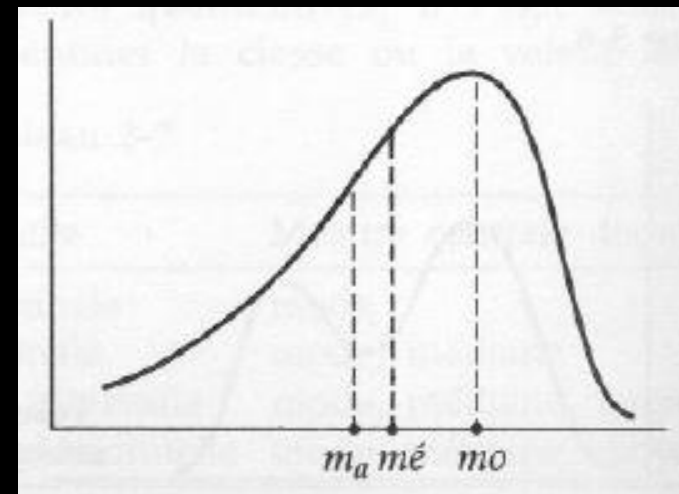
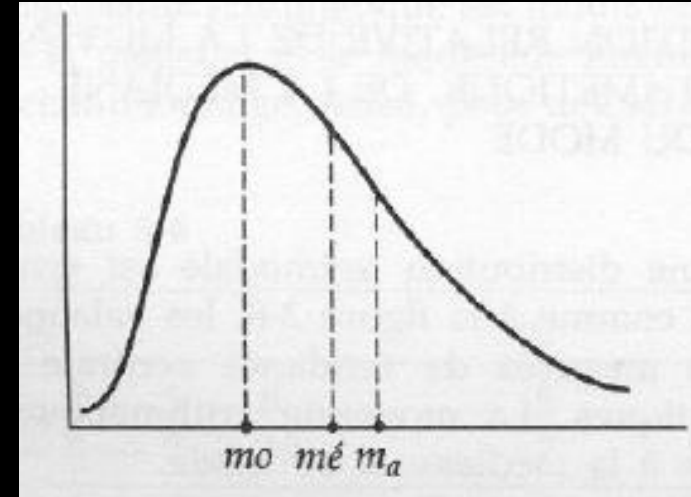


Figure 2.7. Normal (Gaussian) Distribution.



O que fazer?:

- Iniciar a investigação com um grupo (conhecimento do problema) e adicionar competências em “epidemiologia” (desenho) e “bioestatística” (tratamento de dados)
- Desenhar adequadamente o estudo
 - Tipo de estudo (descrição completa e clara)
 - Métodos e técnicas de controlo dos viéses e do confundimento
 - Antecipação do “erro”
- Escolher, definir e “mensurar” as variáveis
- Definir UM (por vezes dois é demais) objectivos/endpoints
- Definir os indicadores (de estrutura, de processo, de resultado)
- Desenhar o plano de amostragem e calcular a amostra (necessária para encontrar..., com que significância, com que margem de erro)
- Estabelecer o plano de tratamento e análise de dados
- AGORA → Recolher os dados e, se o anterior estiver bem feito, será pouco necessário rezar para que tudo corra bem!