



## *Guia de Boas Práticas para Pesquisas Utilizando Forçantes Múltiplas em Ciências do Mar*

---

### **Analisando Dados**

- Tutorial:** O vídeo tutorial [Analisando dados](#) pode ser encontrado no canal do YouTube [MEDDLE para pesquisas com forçantes múltiplas](#).
- Apresentador:** [Peter Dillingham](#), Universidade de Otago, Nova Zelândia
- Vídeo:** [Christina McGraw](#), Universidade de Otago, Nova Zelândia
- Transcripts:** Rebecca Zitoun, Universidade de Otago, Nova Zelândia
- Recursos:** O conjunto de recursos para o *Guia de Boas Práticas para Pesquisas com Forçantes Múltiplas* estão disponíveis no sítio web do [MEDDLE](#).
- 

#### **0:00 - Introdução**

Agora que você assistiu aos vídeos apresentados por [Philip](#) e por [Jon](#) já tem mente conceitos sobre [como montar um] inventário de forçantes para o seu experimento, bem como uma ideia básica [para a construção] de seu desenho. Agora vamos nos aprofundar um pouco mais em desenhos amostrais e pensar um pouco porque eles são importantes.

Um bom desenho experimental lhe dará a capacidade de responder suas perguntas centrais, otimizar a informação, conhecer e ser capaz de defender suas premissas, reduzir custos e ajudar-lhe a alocar melhor recursos limitados, ajudar a publicar seus resultados, e evitar catástrofes.

*Texto (0:20): Um bom desenho experimental irá lhe ajudar: responder suas perguntas científicas, otimizar a informação, conhecer e ser capaz de defender suas premissas, reduzir custos e alocar melhor recursos limitados, publicar, e evitar catástrofes.*

#### **0:41 – Terminologia estatística**

Em primeiro lugar, vamos falar sobre a terminologia que usamos em estatística. Um **FATOR** é uma variável que você planeja manipular. Em um experimento que examina sistemas oceânicos em alteração nós costumamos a nos referir a forçantes como fatores. Então, no experimento de **3 FATORES** apresentado por Jon, ele manipulou a pCO<sub>2</sub>, a salinidade e a temperatura.

*Texto (0:47): Um FATOR é a variável que você pretende manipular. FORÇANTES = FATORES.*

**NÍVEIS** são os diferentes valores que irão ser aplicados a um fator durante um experimento. Por exemplo, a  $pCO_2$  pode ser estabelecida para os valores, portanto **NÍVEIS**, do passado, do presente e do futuro.

*Texto (1:03): NÍVEIS são os valores atribuídos a um fator durante o experimento.*

**TREATMENTOS** são as combinações entre fatores a níveis distintos que são realizados durante o experimento. Por exemplo, quando Jon apresentou seu cenário dos vetores principais, cada frasco para cultivo representou um tratamento.

*Texto (1:13): TREATMENTOS são combinações de diferentes NÍVEIS de FATORES.*

O número de **TREATMENTOS** que você precisa tem um papel muito importante no esforço e nos custos de seu experimento. Então, para realizar um grande número de tratamentos, primeiro você deve ter certeza que realmente tem uma boa razão para isso.

*Texto (1:27): Geralmente, quanto mais TREATMENTOS, maior é o custo e o esforço.*

O que queremos é um experimento simplificado dentro de limites razoáveis, não apenas o mais simples possível.

*Texto (1:42): Desenhe um experimento simples na medida do possível.*

Uma **UNIDADE EXPERIMENTAL** é a unidade primária da análise. Isso pode soar meio vago, porque na verdade é. Você ficaria surpresa(o) ao saber quanto tempo a(o)s estatística(o)s gastam pensando sobre ‘qual é o exato número de amostras ‘n’?’ E isso justifica por que somos a animação da festa.

*Texto (1:51): Uma UNIDADE EXPERIMENTAL é a unidade primária da análise.*

Isso também nos leva a pensar sobre replicações dos diferentes níveis e do problema que [existe] a respeito de replicação *in-situ*. Muitos estudos sobre mudanças climáticas incluem replicações em mais de um nível. Por exemplo, um tanque para cultivos que possui um tanque principal que controla vários [outros] tanques ligados a ele possui dois níveis de replicação. Se isso é ignorado na análise, é aqui que entra a replicação *in-situ*. Se você quiser

saber mais sobre esse assunto, eu indico o trabalho feito por Cornwall e Hurd no ICES Journal of Marine Science.

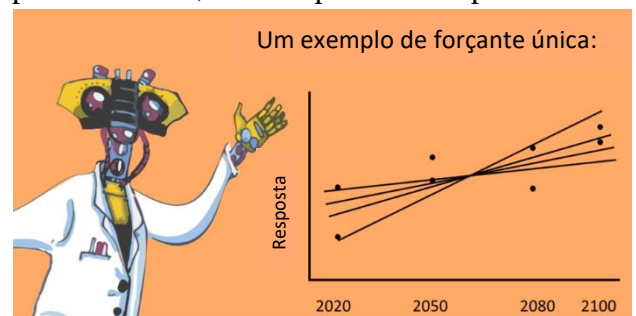
Texto (2:37): [\*ICES Journal of Marine Science\* 73, 572-581 \(2016\)](#)

## 2:46 – *Porque nos importamos?*

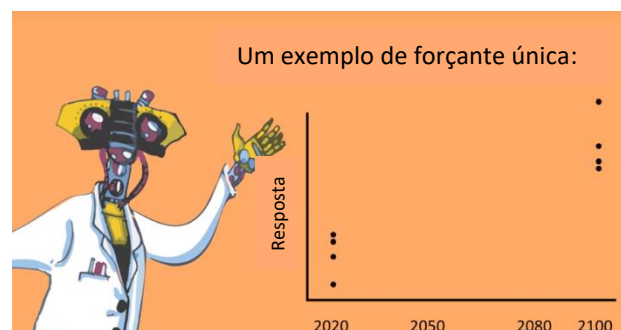
Então, porque nos importamos com tudo isso? Bem, nós queremos desenhar um experimento que minimize o esforço [enquanto] maximize as informações que fornece. Mas nós também precisamos ser realistas sobre o que podemos alcançar.

Texto (2:46): *Porque nos importamos? Nosso objetivo é minimizar o esforço e maximizar as informações.*

Por exemplo, você pode querer explorar resultados com 4 níveis temperatura, representativas digamos, de agora, e do futuro no ano de 2100, além de dois pontos entre eles, para obtermos alguma [noção da forma da] função que a resposta possa ter. Mas, se você pode fazer apenas um número limitado de réplicas, pode resultar num ruído das análises muitas vezes maior que o valor do sinal entre os diferentes tratamentos. Talvez nós fiquemos tentados a traçar uma linha entre os pontos que parece com esta, mas nós poderíamos igualmente desenhar outras curvas, como essas aqui.



Mas digamos que vamos concentrar nossos esforços em apenas dois tratamentos, que representam o presente e o ano de 2100. Isso nos possibilita duas vezes mais réplicas para cada tratamento. Com um número maior de réplicas, nós podemos ser capazes de obter uma resposta um pouco mais útil, enquanto que antes nós não tínhamos informações suficientes para dizer nada proveitoso. Então, temos de ter consciência sobre quanto ruído nós temos, bem como o tipo de sinal que esperamos ter.



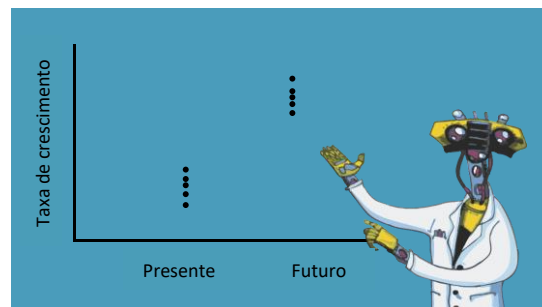
### 3:58 – Desenho 1: O Cenário – cenário com 2 níveis

Vamos levar essas ideias para dentro de um dos experimentos que o Jon apresentou, ‘O Cenário’. Este (*o cenário com 2 níveis*) é um desenho muito bom se nós queremos saber o que acontecerá entre hoje e o futuro, ou entre dois pontos de tempo.

*Texto (3:58): Desenho 1: O Cenário – Jon discutiu um cenário de 5 níveis no tutorial sobre Desenho Experimental. Nós vamos começar com um cenário de 2 níveis.*

É eficiente, mas lembre-se que estamos obtendo pouquíssimas informações fora desses dois pontos. Então, não teremos uma função resposta ou nenhum detalhe como esse, mas seremos capazes de dizer o que está acontecendo agora e durante o futuro. Já que esse desenho (o cenário de 2 níveis) é um arranjo bem simples, ele trás consigo uma boa oportunidade para falarmos um pouco sobre alguns detalhes estatísticos que são também relevantes para os outros estudos. Vamos começar gerando alguns

dados. Nós iremos assumir que estamos olhando para o presente e para o futuro e que estamos medindo as taxas de crescimento para o nosso organismo [*de interesse*]. Digamos que os resultados de nosso experimento parecem



com esses. Vamos também assumir que esses pontos são independentes. Se eles não forem, o que pensamos ser um efeito de agora versus o futuro pode, na verdade, ser apenas um efeito de uma geladeira colocada próxima a janela comparada a uma geladeira colocada próxima a uma porta.

Como analisaríamos esses dados? Primeiro, devemos nos lembrar da pergunta original.

*Texto (5:08): Como analisaríamos esses dados? Primeiro, lembre-se da pergunta original.*

Ela [*a pergunta*] pode ser ‘*Existe diferença entre agora e o futuro?*’ Bem, sim, existe! Se somos espertos sobre qual experimento estamos analisando, nós provavelmente já temos bastante certeza que existe uma diferença entre agora e o futuro. A pergunta mais interessante é ‘*Qual é essa diferença?*’ Nós analisamos vários outros trabalhos já disponíveis antes de começar o nosso experimento, e, temos uma boa ideia de que existirá uma diferença.

*Texto (5:25): Ao invés de perguntar ‘EXISTE diferença?’ seri amais interessante perguntar ‘QUAL é a diferença?’*

Abordagem de teste de hipóteses, i.e., analisando valores de  $p$ , não é o que queremos fazer. Não é muito interessante. E pode nem mesmo ser apropriado. Se você sabe de antemão que a hipótese nula não pode ser verdadeira, então testa-la não é uma boa ideia.

*Texto (5:39): Um teste de hipótese NÃO é apropriado se você já sabe que sua hipótese nula é falsa.*

*Texto (5:53): Aviso de dilema estatístico sobre valores de  $p$  – ‘...o valor de  $p$  ...não pode determinar se uma hipótese é verdadeira ou se os resultados são importantes’ ([Nature 531, 151 \(2016\)](#)).*

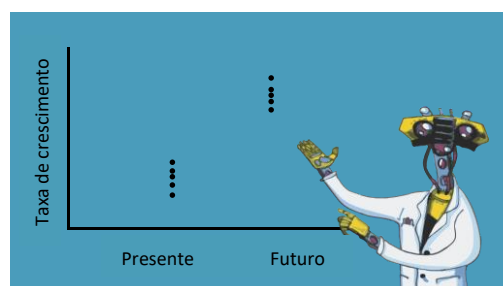
Isso nos leva a algo que tem preocupado profissionais de estatística já a algum tempo, que é o depósito de muito crédito em valores de  $p$ .

*Texto (5:59): Mudando para um mundo além de ‘ $p < 0.05$ ’ – ‘...já é tempo de parar de usar o termo significativamente estatístico completamente. Nem deveriam suas variantes como significativamente diferente,  $p < 0.05$ , e não significativo sobreviver...’ ([The American Statistician 73, S1, 1-19 \(2019\)](#)).*

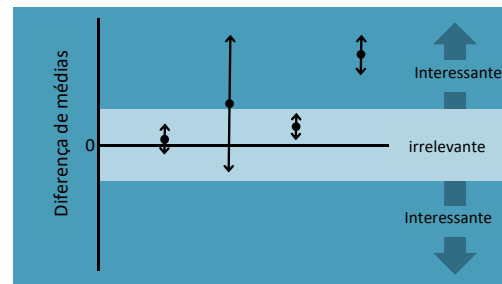
Ao invés disso, estamos mais interessado em estimar e interpretar.

### **6:06 – Estimativas e interpretação**

Assim para comparar o presente versus o futuro, nós podemos pensar sobre estimativas e interpretação, da seguinte forma:



Primeiro nós queremos medir a diferença entre as médias de nossos dois tratamentos. Zero é quando não existe diferença. Nós podemos também pensar em um valor que, apesar de diferente de zero, geralmente poderia ser ignorado em uma perspectiva biológica, e apenas quando os pontos ultrapassarem esse limite, [é que] os resultados começam a ficar interessantes. Então vamos analisar 4 pontos de estimativas, e os intervalos de confiança da diferença, que representam 4 resultados experimentais possíveis. Note que os primeiros dois [resultados] retornariam valores não significativos, se nós fossemos realizar um teste para a hipótese nula, enquanto os dois últimos [resultados] retornariam valores significativos.



*Texto (6:40): Um teste de hipótese nos diria: (a) os dois resultados na esquerda não são significativos, b) os dois resultados na direita são significativos.*

*Texto (6:59): Os intervalos de confiança nos dizem muito, muito mais.*

Mas, intervalos de confiança nos dizem muito, muito mais. O primeiro intervalo nos expõe se a diferença é positiva, negativa ou se ela não existe. Mas apesar disso, o seu efeito não será importante. O segundo intervalo nos informa que provavelmente não desenhamos o nosso estudo muito bem. Nós não sabemos se o nosso resultado é positivo ou negativo, nem se ele [o resultado] é importante ou irrelevante.

Esta é uma boa razão para querermos fazer cálculos de poder estatístico antes de um experimento, ou fazer um estudo piloto se necessário. Mas as vezes, mesmo quando você faz estes [cálculos de poder estatístico], pode não ter muita sorte. Você pode ter colocado estimativas iniciais incorretas em seus cálculos ou pode obter uma base de dados estranha.

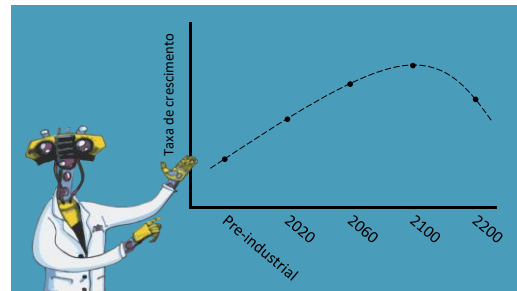
Quando nós chegamos ao terceiro intervalo, nós sabemos que existe um efeito positivo, mas ele é pequeno demais, e assim determinamos que ele não é muito importante. O quarto intervalo é o tipo de coisa que faz as pessoas se animarem. Nós sabemos que o efeito é positivo e sabemos que ele [o efeito] tem uma magnitude importante. Assim, ao invés de simplesmente fazermos um teste-t e perguntar se existe uma diferença, vamos tentar fazer

perguntas melhores, como, ‘Qual é a diferença entre agora e o futuro?’, ‘Como eu vou medir esta diferença?’, ‘Esta diferença é importante?’

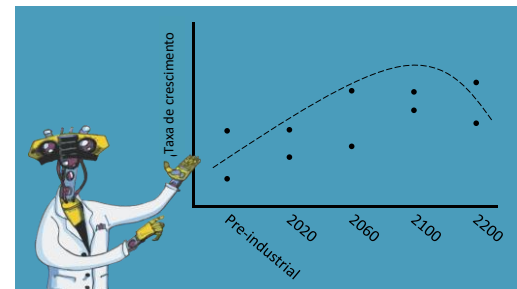
Texto (8:06): ‘Qual é a diferença entre agora e o futuro?’, ‘Como eu vou medir esta diferença?’, ‘Esta diferença é importante?’

### 8:06 – Analisando os dados

Vamos tornar esse cenário um pouco mais interessante. Se pudermos adicionar 3 tratamentos ao nosso cenário, teremos uma chance de observar a forma da curva resposta.



Todavia, a biologia é variável, e dependendo de quanta variabilidade temos e de quantas réplicas temos, nós podemos ou não ser capazes de observar a forma desta curva. Aqui é quando o [simulador do MEDDLE](#) pode nos ajudar de verdade. Tente explorar essa ferramenta e ver se os dados que você obteria irão lhe ajudar a responder a pergunta que você precisa responder.



Você pode ter notado que os valores extremos que nós escolhemos para o nosso cenário de 5 pontos estão fora das condições que nós estamos mais interessados, que seriam agora e o ano de 2100. Isso nos permite testar extremos, o que trás mais poder estatístico geral.

Particularmente, você pode estar procurando por algum ponto de inflexão, um máximo ou um mínimo local que poderia acontecer no futuro mais adiante.

Analisar esta base de dados [do cenário com 5 pontos] pode ser um pouco mais complexo do que para a nossa abordagem de cenários. Uma vez que você adquiriu seus dados, você tem algumas opções. Você pode simplesmente analisar cada tratamento de forma independente como em uma ANOVA. Ou, você pode querer ajustar uma curva resposta, talvez uma reta, talvez uma função quadrática, ou alguma forma de interpolação [alisada] dos valores (ajuste) como uma GAM (modelo generalizado linear), que descreva o seu gradiente.

*Texto (9:13): Analisando sua abordagem com 5 cenários: Analise cada tratamento de forma independente (e.g., ANOVA); Ajuste as respostas a uma curva que descreve o gradiente, e.g., linear, quadrática, GAM.*

Sua abordagem de análise irá depender das perguntas que você quer responder. E também irá determinar de quantas réplicas você precisa e quais premissas terá de fazer. Aqui, mais uma vez, é quando o [simulador do MEDDLE](#) pode lhe ajudar: Você pode tentar ideias diferentes, gerar dados, verificar se pode mesmo analisar estes dados, e conferir se respondem as perguntas que você está tentando responder. Lembre-se, um pouco de tempo que você gastar em simulações pode lhe economizar muito tempo no laboratório.

*Texto (9:58): Simulações são a forma mais fácil de economizar tempo no laboratório.*

### **10:05 – Fatorial completa**

E agora para algo completamente diferente. No outro extremo na escala de complexidade temos o desenho fatorial completo. Fatoriais completos são uma ótima forma de examinar todos os detalhes, quando existem limitações de recursos. Quer dizer, você pode observar os efeitos principais, interações em duas vias, interações em três vias, e assim por diante. Mas elas podem ficar ridiculamente complicadas muito rápido.

Se você fizer uma fatorial completa, ainda poderá fazer projeções entre o presente versus o futuro, todavia, suas projeções não serão mais acuradas do que quando fez o cenário de dois níveis: presente e futuro.

*Texto (10:29): Fatorial completa – Você ainda pode fazer projeções do presente versus o futuro.*

Assim, se você pensa sobre isso, vai perceber que, digamos que você tenha 3 fatores, 2 níveis para cada fator = 8 tratamentos. E se você fizer 5 réplicas por tratamento você necessitará um total de 40 réplicas para uma fatorial completa.

*Texto (10:47): Se você fizer uma fatorial completa com 3 fatores, 2 níveis por fator, terá 8 tratamentos. Com 5 réplicas por tratamento, você precisará de 40 réplicas para executar uma fatorial completa.*

Se você fizer o cenário de 2 níveis, terá o mesmo grau de precisão com apenas 10 réplicas.

*Texto (10:59): Apenas interessado em presente vs. o futuro (i.e., cenário de 2)? Tenha o mesmo nível de precisão com apenas 10 réplicas.*



### **11:02 –Fatorial condensado**

Agora, vamos falar sobre o fatorial condensado. O fatorial condensado é um bom desenho quando você se interessa mais por um fator, dito fator dominante, e por um número de outros fatores que você pode condensar como um único fator composto.

*Texto (11:06): Fatorial condensado: mantenha um fator dominante isoladamente e combine os outros fatores em um único.*

Fatoriais condensados são bons para fazermos projeções e também podem oferecer algum entendimento mecanicista.

*Texto (11:20): Fatoriais condensados são bons para projeções e lhes dará algum entendimento mecanicista.*

Por exemplo, eles podem prover informações [*para descobrirmos*] se por acaso o fator assumido como o dominante, é de fato o dominante, ou, se o efeito dos outros fatores combinados, e as interações entre os mesmos, são importantes, e alguma informação entre a interação do fator dominante e os outros fatores.

*Texto (11:25): Fatoriais condensados podem dizer se o fator assumido como dominante é mesmo dominante. Fatoriais condensados podem dizer se o fator combinado dos outros fatores é importante. Fatoriais condensados podem fornecer informações sobre a interação do fator dominante e os outros fatores.*

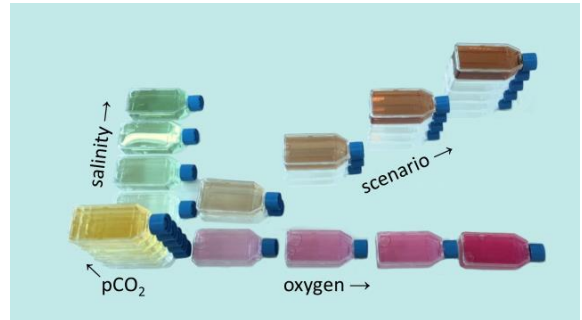
Este será um bom desenho, se você tiver um limite para o número de unidades experimentais que pode fazer [*em cada experimento*], mas você ainda quer obter alguma separação entre o fator dominante e os outros fatores.

*Texto (11:44): Fatoriais condensados são bons se você tem unidades experimentais limitadas, mas ainda quer obter alguma separação entre os efeitos do fator dominante e dos outros fatores.*

Você tem de aceitar que os outros fatores estarão confundidos entre si. Então a informação que obterá é um pouco limitada, porém, você pode ter outras cartas na manga, como a proteômica, que pode ajudar a discriminar alguns dos outros efeitos.

### 12:10 – Fatores Principais

A seguir, falemos sobre vetores principais. Vetores principais permitem que você investigue os fatores individuais bem como seus efeitos combinados. Você possivelmente irá analisar cada um dos diferentes vetores separadamente. Então,



para o exemplo do Jon, nós gostaríamos de montar a análise da seguinte forma: com cada um dos 3 três principais fatores (*pCO<sub>2</sub>*, *oxigênio*, e *salinidade*) para obter curvas de resposta individuais, bem como investigando a diagonal, obtendo um desenho dos cenários.

### 12:10 – Desenho Experimental (ou despenda de maneira esperta)

Uma das principais mensagens que queremos com essa narrativa é dizer que não existe uma única resposta certa para qual desenho experimental usar.

*Texto (12:44): Não existe uma única resposta certa para qual desenho experimental usar.*

Você deve pensar sobre quais são as perguntas que realmente quer responder, e qual desenho irá ser o melhor a ser executado com os recursos que você tem. Por exemplo, como observamos na tabela [a seguir], se você quiser realizar um estudo com uma fatorial completa com 3 fatores, e você precisa de 5 réplicas para ter poder [estatístico] suficiente, com dois níveis por tratamento, você precisaria de 40 unidades experimentais.

*Texto (12:54): Para um estudo de 3 fatores com 5 réplicas por tratamento:*

*Alternativamente, um fatorial condensado com 12 níveis de tratamentos e 12 réplicas necessitará de apenas 48 unidades experimentais.*

Levels per driver	Total number of experimental units			
	Full Factorial	Major vectors	Scenario	Collapsed factorial
2	40	25	10	20
4	320	65	20	20
6	1080	105	30	20
8	2560	145	40	20
10	5000	185	50	20
12	8640	225	60	20

Com um número próximo de unidades experimentais, você poderia utilizar um desenho de cenário com 8 níveis de tratamento e 5 réplicas de cada nível, ou, um desenho fatorial condensado, com 12 níveis de tratamento e 12 réplicas.

Outra forma de olhar para isso seria, digamos, que você possui apenas 24 tanques para cultivo. Então, a não ser que você replique no tempo, estará limitado a esses recursos disponíveis.

*Texto (13:20): Alternativamente, se você possui apenas 24 frascos de cultivo, você pode comparar as possíveis abordagens.*

Number of factors (i.e. drivers)	Replicates per treatment	Levels per factor	Total number of experimental units			
			Full Factorial	Major Vectors	Scenario	Collapsed Factorial
3	2	2	16	10	4	8
		4	128	26	8	8
		6	432	42	12	8
		8	1024	58	16	8
		12	3456	90	24	8
3	3	2	24	15	6	12
		4	192	39	12	12
		6	648	63	18	12
		8	1536	87	24	12
		10	3000	111	30	12
3	4	2	32	20	8	16
		4	256	52	16	16
		6	864	84	24	16
		8	2048	116	32	16
		12	5184	135	36	16
3	6	2	48	30	12	24
		4	384	78	24	24
		6	1296	126	36	24
		8	3072	174	48	24
		12	10368	270	72	24

For the table above: Número de fatores (i.e., forçantes) / Réplicas por tratamento/ Níveis por fator / Número total de unidades experimentais { Fatorial Completa / Vetores Principais / Cenário/ Fatorial Condensada

Lembre-se que você não precisa responder a todas as perguntas interessantes com um único experimento. É muito melhor responder bem a uma pergunta do que a várias perguntas não tão bem.

### 13:43 – Dicas e Truques – ou como eu passo as minhas sextas-feiras a noite.

Você está aqui provavelmente porque tem um pergunta realmente interessante que pretende responder. Antecipa que será complexo e difícil de responder. E pode necessitar investir bastante tempo. Então, quanto mais você se planejar, e antecipar, terá melhores garantias de ficar feliz com o resultado final. Coisas que eu recomendo são:

- Saiba como analisar os seus dados antes de iniciar o seu experimento.
- Se você tem informações suficientes para fazer um cálculo de poder estatístico, você deve fazê-lo.
- Simule dados e analise seus dados simulados. Lembre-se que os dados reais quase nunca serão tão bons como seus dados simulados. Se você não puder analisar seus dados simulados para responder a pergunta que quer fazer, você estará com grandes problemas posteriormente, quando obtiver seus dados reais.
- Use os resultados de suas simulações como uma oportunidade para refinar sua pergunta e redesenhar seu experimento.
- Pequenos estudos piloto são seus amigos. Você vai aprender como os dados se comportam. Por exemplo, vão mostrar se você deve trabalhar com uma escala

linear ou logarítmica. Eles [*estudos piloto*] também fornecerão uma ideia melhor do nível de ruído que você deve esperar, bem como de quantas réplicas você precisa ter.

- Para desenhos que não tem muitos níveis de tratamento, é sempre bom ter réplicas extra. As vezes, coisas estranhas acontecem, e você pode perder uma de suas culturas, ou, você pode descobrir algo realmente interessante.
- Se você tem a necessidade, ou a oportunidade, de consultar um profissional de estatística, ou um(a) cientista mais experiente que saiba um pouco mais de estatística que você, então você deve fazer isso antes do início de seu experimento.

Por fim, tenha coragem e não tenha medo de tentar coisas novas. Isso irá tornar seu experimento realmente interessante para todos.



GORDON AND BETTY  
**MOORE**  
FOUNDATION



ANTARCTIC CLIMATE & ECOSYSTEMS  
COOPERATIVE RESEARCH CENTRE

