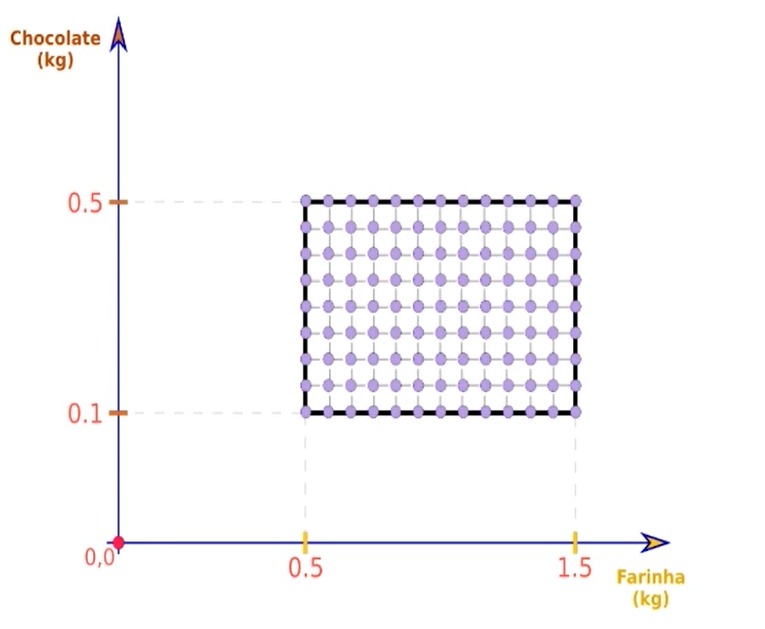
Anteriormente aprendermos a fazer com que o processo de coleta de dados seja finito. Nosso objetivo agora é analisar algumas propostas de escolha de quais prontos serão coletados para o experimento, de modo que, ao final, consigamos escolher aquelas que melhor satisfaçam as necessidades do projeto.

Nossa primeira proposta será, a partir da região que estabelecemos anteriormente, escolher uma linha e uma coluna, realizando o experimento nessas condições.



Vamos imaginar que escolhemos a linha central da nossa área de experimentação, ou seja, a quinta linha, usando apenas os pontos que a compõem para realizarmos o experimento. Analisando esses pontos, é possível perceber que estamos variando a quantidade de farinha, o que implica em recolher dados que representam o impacto deste ingrediente na quantidade de cupcakes produzidos.

Entretanto, nos deparamos com um problema: nós não variamos o chocolate, que permanece fixo em todos os pontos. Assim, atendemos apenas parcialmente aos objetivos do experimento, já que não avaliam o efeito do chocolate na quantidade de cupcakes produzidos.

Você pode se perguntar se não é possível tirar essas informações na fase de análise estatística, e a resposta é não. Afinal, se nossos dados experimentais não incluem determinado fenômeno, nenhum método estatístico ou análise matemática, por mais rebuscados que sejam, serão capazes de extrair essas informações.

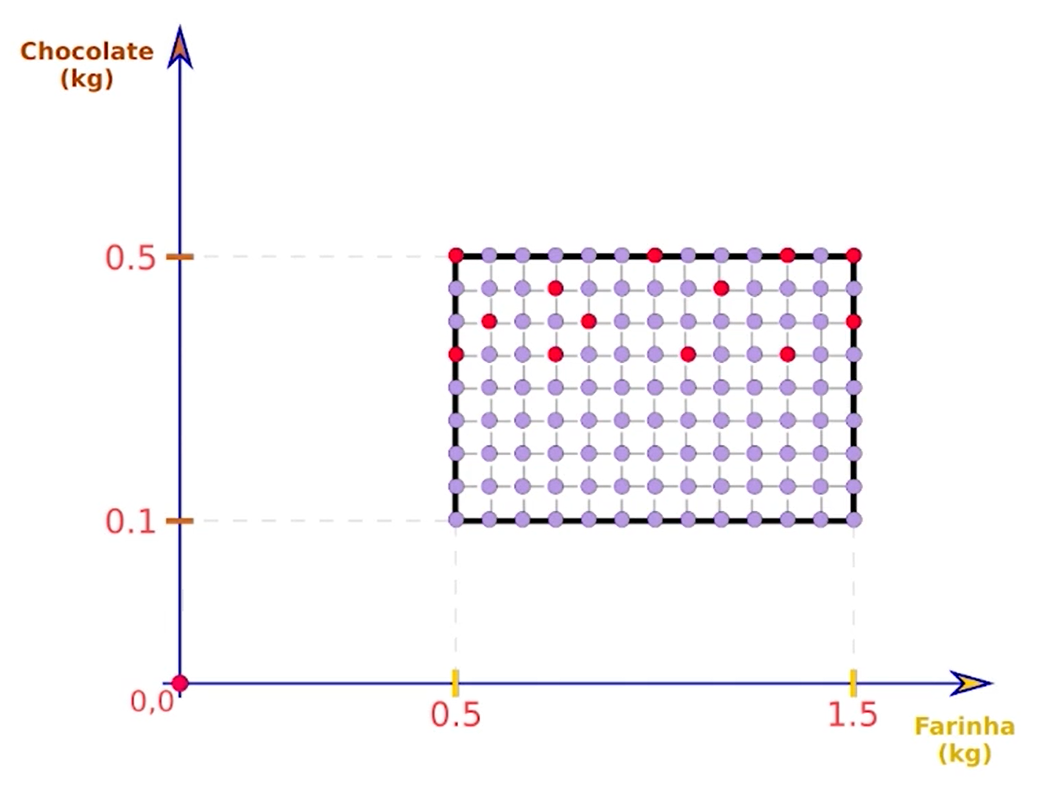
Sabendo que precisamos variar tanto a farinha quanto o chocolate, chegamos a uma segunda opção: variar um elemento de cada vez. Por exemplo, na opção anterior tínhamos fixado o chocolate e variado a farinha em toda a sua extensão. Em seguida, poderíamos fixar a farinha numa posição e variar somente a quantidade de chocolate em toda a sua extensão, por exemplo escolhendo a coluna central da região estabelecida (a sétima coluna).

A princípio parece que estamos atendendo ao objetivo do experimento, já que avaliaremos o efeito das duas variáveis manipuláveis dentre os limites inferiores e superiores. Entretanto, o fato de variarmos um elemento de cada vez é na verdade um problema, pois não captaremos a interação entre eles.

Os pontos referentes à farinha, por exemplo, equivalem a uma quantidade fixa de chocolate. Porém, é possível que a variação da quantidade de chocolate tenha um efeito diferente sobre os experimentos. Sendo assim, por mais que essa opção pareça mais simples e lógica para as análises estatísticas, ela não nos possibilita verificar o efeito de interação entre as variáveis manipuláveis.

Apesar de contra intuitiva, a melhor forma de realizarmos o experimento é variando em conjunto ambas as variáveis manipuláveis. Para isso, escolheremos os pontos de forma aleatória. Começaremos definindo uma número de dados experimentais, e depois, a partir de um procedimento aleatório, escolheremos os dados a serem coletados.

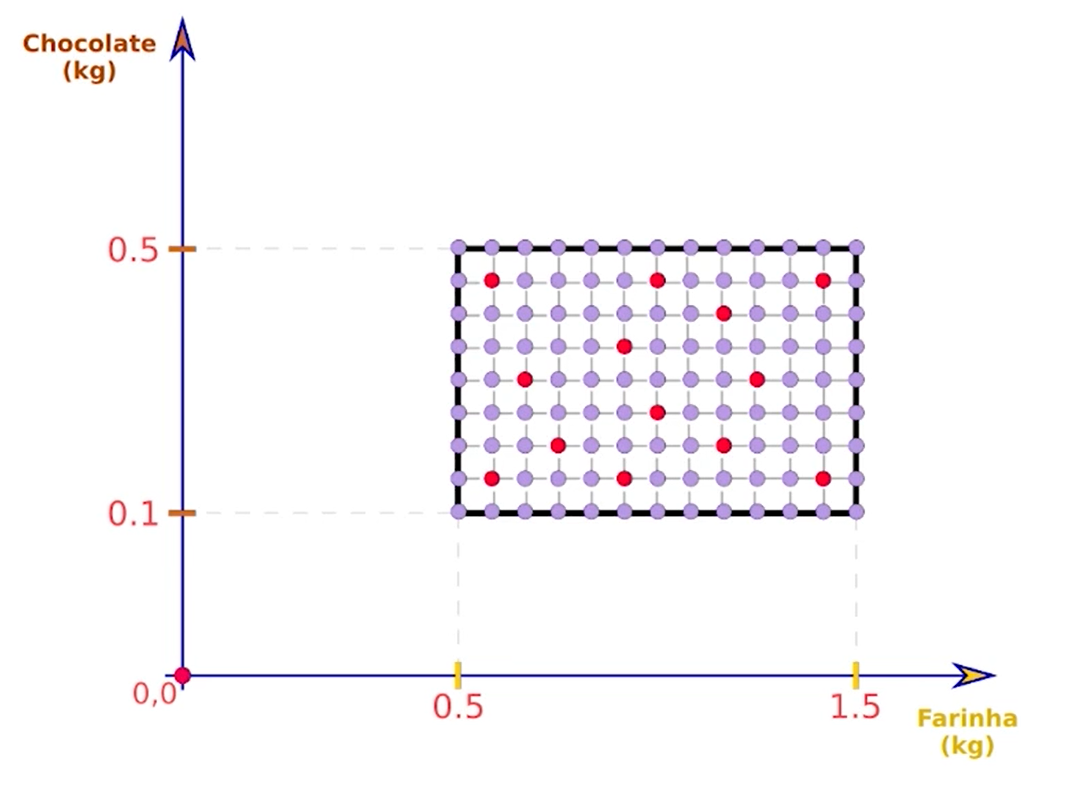
Digamos que um procedimento aleatório escolheu aleatoriamente 13 pontos, que representaremos abaixo.



A princípio, é verificável que tais pontos incluem a variação de ambos os ingredientes e estão captando efeitos de interação entre eles, já que temos diferentes quantidades de farinha e chocolate. Contudo, os pontos experimentais escolhidos aleatoriamente, nessa conformação, estão todos reunidos na parte superior da área experimental, sem nenhum ponto na parte inferior. Dessa forma, não temos nenhuma informação sobre como se comporta nossa receita para tal parte da área experimental, ainda que tenhamos melhores resultados para outros critérios.

Em outras palavras, essa configuração dos pontos não é representativa de toda a área experimental, e precisamos que os pontos representem toda a região de estudo.

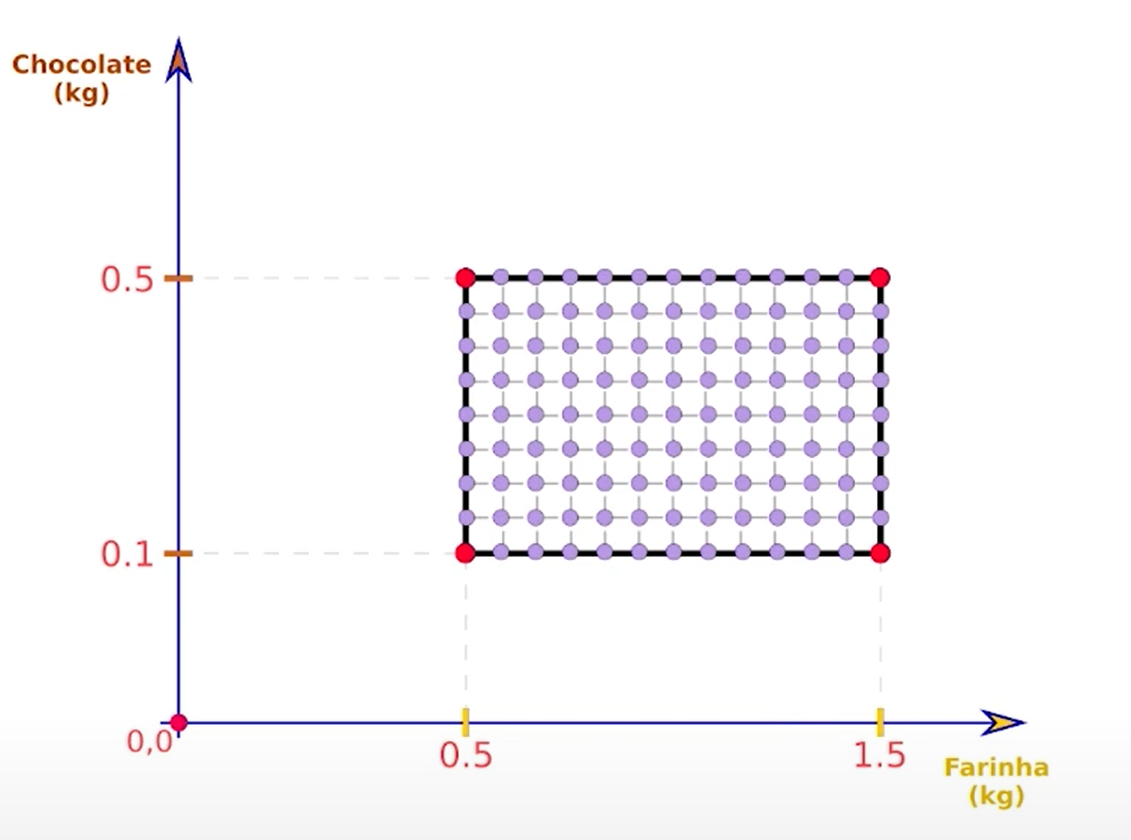
É possível afirmar que a distribuição que privilegia a parte superior seja uma casualidade, e que outro conjunto de números aleatórios possa escolher uma configuração mais interessante. Pensando nisso, digamos que nosso procedimento aleatório foi executado novamente, escolhendo 13 novos pontos para realizarmos o experimento.



Esses pontos parecem estar distribuídos de maneira mais homogênea na área experimental, além de atingirem os outros objetivos. Contudo, esses pontos ainda não incluem os limites inferiores ou superiores da quantidade de chocolate ou de farinha. Nesse caso, poderíamos aumentar o número de dados experimentais, garantindo que percorremos a área experimental em toda a sua totalidade.

Entretanto, não querermos gerar dados experimentais em excesso, afinal cada ponto experimental é uma receita de cupcake que a Bel precisará preparar, o que implica em um custo. Isso sem contar que estamos lidando com um cenário bastante simples, e, caso tivéssemos mais variáveis manipuláveis, seria ainda mais difícil enxergar se um procedimento aleatório está deixando de representar alguma área experimental. Isso significa que confiar muito em um procedimento aleatório pode não ser a melhor saída.

Tendo apresentado essas opções, vamos introduzir uma nova abordagem, bastante interessante quando a coleta de dados é custosa: o planejamento fatorial. Essa abordagem consiste em analisar os pontos nos vértices da área experimental. Agindo assim, o conjunto de pontos representará todas as possíveis combinações entre o limite inferior e superior das variáveis manipuláveis.



Analisando a conformação desses pontos, é possível perceber que eles capturam toda a variação de quantidade de farinha e chocolate - ou seja, são representativos de toda a área experimental -, e ao mesmo tempo captam efeitos de interação entre a quantidade desses ingredientes.

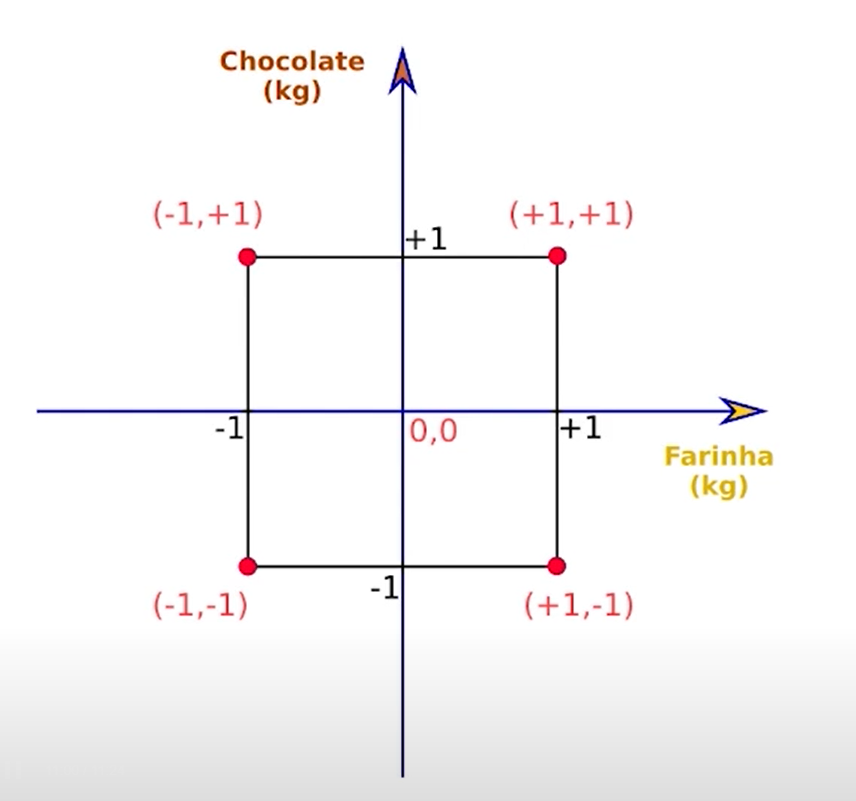
Outra questão interessante do planejamento fatorial é que conseguiremos atingir esses objetivos realizando poucos - nesse caso, apenas quatro. Para sabermos quantos experimentos serão realizados no planejamento fatorial, basta elevarmos o número 2, que representa os níveis (inferior e superior), por um número v que representa a quantidade de variáveis manipuláveis. Ou seja, como temos dois níveis e duas variáveis manipuláveis, nosso cálculo será 2^2 (dois elevado a segunda potência), cujo resultado são 4 ensaios.

Pegaremos esses dados e os incluiremos uma tabela na qual cada linha representa um dos ensaios do experimento da Bel.

| **Ensaio** | **Farinha (kg)** | **Chocolate (kg)** | **Porções** |
| --- | --- | --- | --- |
| 1 | 0.5 | 0.1 |  |
| 2 | 1.5 | 0.1 |  |
| 3 | 0.5 | 0.5 |  |
| 4 | 1.5 | 0.5 |  |

É tradicional na realização de planejamentos fatoriais que os pontos experimentais sejam normalizados entre -1 e +1. Para isso, deslocaremos a nossa área experimental no sentido da origem do plano cartesiano. Dessa forma, o limite inferior receberá o valor -1, e o superior +1.

Agindo dessa forma, evitaremos problemas que podem surgir com variáveis manipuláveis cujas faixas de variação possuam magnitudes diferentes, ou mesmo que apresentem escalas diferentes, além de podermos incluir variáveis qualitativas no planejamento - por exemplo, as marcas de fermento utilizadas.



Agora que organizamos nossos dados experimentais, realizaremos os experimentos da Bel usando o planejamento fatorial, pois ele atende os objetivos discutidos anteriormente e possibilita que eles sejam atingidos realizando um pequeno número de ensaios.