

07 Observações Sistemáticas e Inferência

Anotações - Observação Sistemática e Inferência

Hartmut Günther

O elemento essencial da pesquisa científica, e, mais especificamente, pesquisa empírica, é o de fazer *observações sistemáticas*. As diferentes maneiras de observar e registrar comportamento humano representam as diversas técnicas de pesquisa, tais como experimento de laboratório, estudo de campo, entrevista, entre outras.

Independentemente de qual técnica utilizada, a opção de realizar observações sistemáticas pressupõe de que *há ordem* subjacente aos fenômenos que estão sendo observados, e ainda, que se trata de uma ordem que é acessível, direta ou indiretamente (i.e., por meio de instrumentos apropriados) a nossas percepções. Em outras palavras, o objetivo do processo científico é o de descobrir e elucidar a ordem dos fenômenos, dentro de um campo de interesse específico.

Tendo coletado dados de maneira sistemática (seja que for a técnica específica utilizada) e sistematizado os mesmos (seja qualitativa ou quantitativamente) o pesquisador defronta-se com a seguinte interrogação diante das suas, necessariamente limitadas, observações: Até que ponto, é possível de estender os achados para além dos dados concretos? Em outras palavras, é possível generalizar, e, caso sim, para que população? Será que os achados representam uma realidade subjacente, ou será que a constelação dos dados encontrada é um simples acidente? Mais formalmente, a pergunta é: *Qual a probabilidade de poder ter obtido aquele conjunto de dados / resultados por acaso?* Embora esta questão seja mais facilmente respondido com base de dados quantitativos, o raciocínio apresentado a seguir, aplica-se a qualquer tipo de análise.

A maneira de responder a esta pergunta, é partir da suposição de que *não existe ordem*. Tentar-se-á, então, demonstrar que tal afirmação é *falsa*. Ao invés de provar a existência de algum fenômeno, tenta-se demonstrar a *inexistência do oposto* àquilo que nos interessa. Supormos que quer-se-ia estabelecer uma relação entre dois eventos, digamos o tamanho do pé direito e a inteligência do seu portador. Ao invés de tentar provar a existência desta relação, tenta-se verificar a *inexistência do oposto*, i.e., inexistência da ausência desta relação.

Neste processo, contrasta-se duas realidades: Como os fenômenos estudados de fato são (ordem real) e como eles parecem na base dos dados coletados (ordem achada). Parte-se da hipótese de que não existe tal

relação (no caso, não existe relação entre inteligência e tamanho do pé direito, razão pela qual é chamada de hipótese nula (H_0)). Realiza-se observações, coleta-se dados que se considera capazes em *falsificar* esta hipótese nula. Na medida em que não é possível provar a inexistência de algo, há apenas a possibilidade de uma aproximação, i.e., uma probabilidade da sua inexistência. Daí a pergunta, *Qual a probabilidade de obter aquele conjunto de dados / resultados por acaso?* Aquele conjunto de dados é que se usa para demonstrar a inexistência da hipótese nula. Caso a probabilidade de se obter estes dados por acaso for suficientemente pequeno, *rejeita-se a hipótese nula* e supõe-se a existência do fenômeno de interesse, no caso, a relação entre tamanho do pé direito e a inteligência. Ao contrário, diz-se *não ter sido possível rejeitar a hipótese nula*, continua-se supor que não existe tal fenômeno, no caso, tal relação. O que, entretanto, quer dizer 'suficientemente pequeno'? É a chance de erro que admitimos no sentido de supor uma existência de algum fenômeno, quando de fato não existe. Na Figura 1, resume-se as quatro possíveis situações.

- I O fenômeno sob estudo (relação entre tamanho do pé e inteligência) de fato existe na ordem real (H_0 é falso). Na medida em que os dados achados sugeriram a existência de tal fenômeno (rejeitando-se H_0), 'acertou-se' a resposta certa. A capacidade de um teste estatístico conseguir levar o pesquisador a esta conclusão se chama *power* de um teste.
- II Note-se que não se tem acesso a ordem real, somente podemos fazer inferências acerca dela.
- III O fenômeno sob estudo de fato não existe na ordem real (H_0 é verdadeiro), mas os dados achados sugeriram a existência de tal fenômeno (rejeita-se H_0). Cometemos o *erro tipo I* ou erro α . O erro α representa justamente a probabilidade de os resultados terem sido obtidos por acaso, portanto não refletindo a realidade. *Erro Tipo I, α* : Supor que existe 'ordem' onde de fato não há. Ordem, neste contexto quer dizer, postular a existência de uma relação entre X e Y, ou de uma diferença entre A e B. Em outras palavras, embora a 'ordem' achada de fato não exista na realidade, *por acaso* aquela constelação de dados foi achada. A probabilidade α representa aquele limite de probabilidade que, ultrapassada, significa que uma ordem achada não é mais considerada por acaso.
- IV O fenômeno sob estudo (relação entre tamanho do pé e inteligência) de fato existe (H_0 é falso).

Os dados coletados, entretanto, não sugeriram tal existência (não consegue-se rejeitar H_0). Cometemos o *erro tipo II* ou erro β . As exigências para se evitar um erro tipo I eram rígidas demais, ao ponto de se cometer este erro tipo II, o de negar a existência de um fenômeno, quando de fato existe.

Error Tipo II, β : Por outro lado, para evitar um erro tipo I, o pesquisador estipula uma probabilidade α muito exigente. Desta maneira pode cometer um outro tipo de erro, o Error Tipo II: Não admitir a existência de uma determinada ordem, quando na realidade ela existe. Quanto menor a probabilidade α , maior a probabilidade β , a possibilidade de cometer um erro do tipo II: Quanto menor α , mais resultados podem parecer ser por acaso, enquanto de fato se trata de uma ordem sistemática.

Ñ O fenômeno de interesse não existe (H_0 é verdadeiro) e os dados coletados também sugerem a ausência de tal fenômeno (não se consegue rejeitar H_0). A conclusão foi acertada.

Resumindo o raciocínio até aí:

- 1 Supõe-se ordem no universo.
- 2 Faz-se observações sistemáticas, dentro de certas condições controladas, i.e., para diminuir o tamanho do 'erro' = variáveis não consideradas.
- 3 Tendo coletados os dados por meio de um processo de observação sistemática e apropriada à pergunta da pesquisa, determina-se, se o conjunto de dados assim alcançado, não poderia ter sido obtido por acaso.
- 4 [Por meio de um teste estatística,] Procura-se resolver esta pergunta.
- 5 Caso a probabilidade de obtenção destes mesmos resultados por acaso se mostra 'suficientemente' baixa, constata-se, que os resultados são

sistemáticos, i.e., refletem a ordem que se supõe existir no universo.

- 6 Se, por outro lado, a probabilidade de obter um 'mesmo' conjunto de dados / resultados por acaso se mostra 'relativamente' alta, conclui-se que os resultados poderiam ser tanto sistemáticos, quanto por acaso, e, por via das dúvidas, não os aceita como refletir um hipotetizada ordem.

Obviamente, não há certeza que a *ordem achada* por meio deste processo de fato corresponda a *ordem real*, no caso do passo 5. Igualmente, não há certeza de que a *ordem rejeitada* por meio deste processo de fato não existe, no caso do passo 6.

Figura 1

¹ Uma relação entre tamanho do pé e inteligência ...

		² ... na ordem real ...	
		³ ... existe H_0 é falso	³ ... não existe H_0 é verdadeiro
² ... na ordem achada ...	³ ... existe (rejeita-se H_0)	Ī $1 - \beta$ Conclusão acertada	İ α erro tipo I
	³ ... não existe (Não se consegue rejeitar a H_0)	Đ β erro tipo II	Ñ $1 - \alpha$ Conclusão acertada