

Estatística na Sociedade

-- uma digressão ilustrativa por domínios de aplicação¹

Carlos Daniel Paulino, Dep. Matemática, IST, UTL e Marília Antunes, DEIO, FCUL

Estatística e o seu papel político-cultural

A informação estatística proveniente de censos, inquéritos amostrais e outras fontes e o pensamento estatístico constituem conhecimento e meio indispensáveis à tomada de decisões sobre a estratégia e políticas de governação dos países e das suas instituições e empresas e à assunção de uma cidadania plena e responsável por parte dos seus habitantes.

A Estatística tem um papel decisivo, designadamente, no planeamento da recolha e na devida organização, tratamento e análise daquela informação. Inclusivamente, a Estatística dispõe de métodos para averiguar se na informação recolhida há suspeita de erros ou de dados forjados que, a existir, deve obrigar a uma investigação adicional. Isto é particularmente relevante em atividades de contabilidade e auditoria visando a deteção de eventuais anomalias e fraudes em documentos comerciais e financeiros.

Estatística nas Ciências da Vida

Biomedicina: campo de estudos na interface entre as ciências da *Biologia* e *Medicina* voltados para a investigação de doenças humanas no que concerne a causas, mecanismos, prevenção, diagnóstico e tratamento. Muitos desses estudos envolvem experiências que precisam de ser coerentemente delineadas e cujos dados exigem análises por metodologias estatísticas adequadas.

I. Estudo de potenciais fatores de risco da infeção no colo do útero pelos vírus do papiloma humano e de comportamentos de risco da infeção pelo vírus da imunodeficiência humana.

II. Estudo da diversidade do repertório dos recetores de linfócitos T (TCR), de que depende crucialmente o sistema imunitário adaptativo, e sua comparação entre os tipos CD4+ ativador e regulador da resposta imunitária (*Imunologia*).

Farmacometria: ramo da *Farmacologia/Biofarmacologia* Quantitativa, a que se dedica a indústria farmacêutica/biofarmacêutica envolvida em investigação, que se ocupa do desenvolvimento e aplicação de métodos estatísticos e matemáticos para caracterizar através de ensaios clínicos a *farmacocinética*, *farmacodinâmica* e resposta de pacientes de doenças biologicamente estudadas à administração de fármacos ou biofármacos (medicamentos produzidos por processos biotecnológicos).

Por outro lado, o desenvolvimento de uma nova droga requer um estudo de degradação da sua potência com o tempo decorrido após o fabrico. Para o efeito, escolhem-se de forma adequada várias amostras da droga e determina-se a degradação da sua qualidade (redução

¹ Este trabalho foi financiado por Fundos Nacionais através da Fundação para a Ciência e a Tecnologia (FCT) no âmbito do projecto PEst-OE/MAT/UI0006/2011.

da sua potência original) ao fim de sucessivos intervalos de tempo, ao longo de períodos pré-especificados. A adoção de modelos estatísticos apropriados para tais dados permite estudar como varia a fiabilidade da droga (probabilidade de se manter suficientemente eficaz) com o tempo e estipular qual o período de validade a recomendar.

Biometria: campo de estudo de modelos e métodos estatísticos e matemáticos aplicados a problemas de análise de dados das *Biociências* (Ciências da Vida):

I. Análise da eficácia relativa de competidoras terapias de doenças através de ensaios clínicos humanos (*Biomedicina, Biofarmacologia*) -- como exemplo, refere-se a comparação de pacientes tratados a cancro do colon por uma de duas terapias atribuídas aleatoriamente com respeito ao tempo desde ingresso no ensaio até à ocorrência de recorrência da doença.

II. Comparação de rendimentos de diferentes variedades de um cereal, ou de produção de distintas raças de uma espécie de gado, ou de eficácia de técnicas de restauração dentária, através de ensaios agrícolas/animais/humanos previamente delineados (**Agropecuária, Medicina Veterinária, Agronomia, Zootecnia, Odontologia**);

III. Análise de efeitos de poluição atmosférica ou aquática na ocorrência de doenças nos residentes de uma dada região (*Ciência do Ambiente, Saúde Pública*);

IV. Estudo de efeitos da prevalência e intensidade de fogos florestais, de alterações climáticas ou de transformações humanas do solo na biodiversidade da região em análise (**Silvicultura, Biologia Vegetal/Animal**).

V. Estudos sobre a identificação de causas de doenças (**Genética**)

A. Os avanços científicos que permitiram a descodificação dos genomas das espécies trouxeram consigo uma mudança de paradigma na análise de dados de natureza biológica. No passado, a situação comum era a de observação de um número pequeno de variáveis sobre um número razoavelmente grande de indivíduos. As atuais tecnologias permitem a obtenção simultânea do nível de expressão de milhares de genes (as variáveis) mas, condicionados pelos elevados custos de produção, os investigadores têm, em regra, possibilidade de obter estes dados para um número reduzido de réplicas biológicas.

Dispondo de dados concretizados em matrizes de elevada dimensionalidade contendo níveis de expressão genética, é possível identificar estatisticamente genes com expressão diferenciada em casos de doença – genes que, como resultado da presença de doença, têm a sua função alterada ao ponto de se verificar um aumento ou diminuição anormal do nível de atividade.

B. Em muitos problemas os fenótipos de interesse exprimem-se de forma dicotómica, por exemplo através de suscetibilidade ou resistência a certas doenças do foro genético. A localização por etapas de regiões cromossómicas associadas com tais fenótipos é um assunto de interesse. Após uma primeira localização de cromossomas suspeitos, o recurso a um escrutínio mais refinado, denominado mapeamento genético por intervalos, usando distintos modelos estatísticos permitiu identificar mais precisamente os locais do genoma associados com a suscetibilidade à malária cerebral em cobaias.

Estatística em Epidemiologia e Saúde Pública

I. Estabelecimento de políticas de saúde pública

O registo de novos casos de doenças juntamente com a sua localização no espaço e no tempo permite um mapeamento espacial/temporal da taxa de incidência dessas doenças e a sua análise estatística pode estabelecer quais as ações a empreender no quadro de defesa da saúde pública.

Estatística em Ecologia e Ciências do Ambiente

I. Estudos de implicações na *Biogeografia* e *Biologia da conservação*

A. Dados: registo de presença/ausência ou de níveis ordinais de abundâncias de espécies em sítios amostrados num conjunto selecionado de zonas da região a analisar, bem como de valores de variáveis ambientais (e.g., temperatura, humidade, tipo de solo, etc.).

Objetivos: Análise através de modelos (espaço-temporais) apropriados para obter respostas a questões do género: Como é que populações de animais e/ou de plantas se distribuem espacialmente e/ou temporalmente? Áreas de alta biodiversidade tendem a causar menor abundância de espécies? Existem áreas ricas quer em biodiversidade quer em abundância, propiciando regiões ideais para a aplicação de esforços de conservação? Há evidência de duas ou mais espécies ocorrerem em áreas geograficamente sobrepostas (*simpatria*) ou isoladas uma da outra (*alopatria*)? Qual o efeito de transformações no terreno (para fins de agricultura, pecuária, silvicultura) ou de variações climáticas na distribuição da respetiva abundância e alterações do habitat?

B. Dados: registo de localizações com presença de dadas espécies no espaço, acompanhado de informação sobre fatores ambientais.

Objetivos: Ficar a saber como distintos fatores ambientais e tipos de solo incentivam ou desincentivam a presença de espécies ao longo de uma região; prever a distribuição da ocorrência destas em regiões não amostradas com conhecidas condições ambientais e possibilitar o delineamento e a gestão de estratégias de conservação. A extensão em que se verifica sobreposição do habitat de várias espécies na região inspecionada constitui uma medida da alta ou baixa diversidade (grau de simpatria) ao longo dessa região, para além de poder evidenciar uma tendência para ausência ou presença de competição no consumo de recursos do solo ao ponto de suscitar a exclusão de umas por outras.

II. Estudo de monitorização ambiental com implicações em saúde pública e qualidade de vida em geral

C. Dados sobre níveis de poluição atmosférica em grandes meios urbanos/industriais

Objetivos: Análise estatística visando averiguar as suas causas e efeitos na saúde dos residentes e implementação de potenciais medidas atenuadoras ou reparadoras (e.g., restrição do tráfego automóvel, filtragem de agentes poluidores, deslocalização de fábricas, etc.).

III. Estudo de previsão de eventos hidrológicos mais ou menos raros (**Hidrologia**)

D. Dados sobre registos históricos de eventos com graves consequências sócio-económicas (cheias, desabamentos, secas) e medidas associadas (quantidade de precipitação, caudais de rios).

Objetivos: Estimar o intervalo médio de recorrência desses eventos e, com base em modelos estatísticos, avaliar parâmetros vitais para análise de risco que influencie tomada de decisões sobre a adoção de medidas preventivas e construção de infraestruturas com limites associados (diques, barragens, valas).

IV. Estudo de evolução de ecossistemas marinhos e atributos físicos oceânicos (**Oceanografia, Geografia Física – Climatologia**)

E. Dados sobre registos passados de temperatura e salinidade da água do mar, velocidade do vento e capturas de pescado.

Objetivos: Averiguar se reduções drásticas em estoques de certas espécies piscícolas têm que ver com alterações climáticas em adição a eventuais esforços pesqueiros exagerados, e se aquelas alterações se devem a fatores naturais ou humanos.

Estatística em Ciências Físicas e Químicas

I. Estudo de separação de sinais de fundo/fontes em imagens na **Astrofísica**

Dados: Imagens astronómicas obtidas por telescópios consistindo de um fundo difuso (resultante de uma composição de emissões instrumentais e cósmicas com flutuações graduais ou abruptas da sua intensidade) no qual estão sobrepostos objetos celestiais de interesse (fontes) com morfologia e luminosidade grandemente variáveis (e.g., estrelas, galáxias em conglomerados ou não, nebulosas).

Objetivos: Detecção e caracterização de fontes indistintas de tipo pontual ou estendido com estimação do fundo através de modelos estatísticos que atendam ao ruído corruptor dos sinais presentes nas imagens e à informação *a priori* acumulada.

Este tipo de pesquisas é relevante para tentar responder a questões fundamentais da Astrofísica tais como a origem do universo e a distribuição da matéria no universo.

II. Estudos relacionados com a **Física e Química Nuclear**

As aplicações da **Radioquímica** (química dos material radioativos) são hoje extremamente variadas. Referem-se os setores da *Energia* com a produção de energia nuclear, das *Ciências Médicas* e *Farmacêuticas* com o diagnóstico e terapia de doenças e da *Arqueologia* com a datação de restos de matéria orgânica pelo Carbono-14 (radioisótopo do Carbono-12).

Sendo a radioatividade um fenómeno tipicamente aleatório, compreende-se então a relevância do tratamento estatístico da desintegração de materiais radioativos, através nomeadamente de registo temporal de contagens de emissões de radiação por apropriados detetores.

III. Estudo de determinação indireta de composição química, do foro da **Quimiometria**

A Quimiometria é o campo da Química consistindo na aplicação de métodos estatísticos e matemáticos para o delineamento otimizado de experiências químicas e extração de informação relevante contida em dados dessas experiências.

Dados: Medição de quantidades obtidas instrumentalmente e relacionadas com a concentração de um dado composto químico na amostra em análise, bem como em várias outras amostras mas agora com composição conhecida.

Objetivos: Predizer a concentração do composto na amostra de interesse através da pesquisa de um modelo estatístico que se revele adequado no ajustamento dos dados obtidos.

Estatística em Engenharia

Em Engenharia é comum o delineamento de processos tendo em vista a avaliação de unidades experimentais (objetos físicos, formulações químicas, estruturas, materiais) quando sujeitas a algum tipo de intervenção (tratamento), especificado por uma ou mais variáveis controladas (fatores).

Os objetivos do planeamento experimental são a melhoria do desempenho e a redução da variabilidade dos processos, o desenvolvimento de novos processos e a redução do tempo e custo da operação global. O uso da Estatística no planeamento propriamente dito das experiências e na análise dos respetivos resultados permite uma maior eficiência e economia, bem como uma objetividade científica na extração das conclusões do processo experimental.

I. Investigação na resistência à compressão de betão (material composto de cimento, como aglomerante, de gravilha e areia como agregado e água) dos efeitos de distintos métodos de cura (processos de impedir a rápida evaporação da água na pasta de cimento), em vários espécimens de betão (**Eng. Civil**).

II. Estudo da evolução de danos estruturais (comprimento de rachas) provocados por fadiga do material em estudo quando sujeito a ensaios sucessivos de cargas cíclicas, tendo em vista a estimação da durabilidade residual medida pela fiabilidade, i.e., probabilidade de o tamanho da racha resistir à fratura como função do nº de ciclos de carga (**Eng. Mecânica**).

III. Averiguação do comportamento da resistência à tração de ligas metálicas (**Eng. Materiais**) ou de fibras sintéticas (**Eng. Têxtil**) em função da sua composição química ou percentagem de algodão, respetivamente.

IV. Estudo do efeito na luminosidade de válvulas integradas em televisores, do tipo de vidro e do fósforo nelas usados (**Eng. Eletrotécnica**).

V. Estudo do efeito no rendimento de um dado processo químico de fatores como o volume de solvente, a temperatura e o tempo de condensação e a quantidade de cada um dos dois materiais envolvidos no processo (**Eng. Química**).

VI. Estudo do volume de tráfego automóvel ao longo de dias úteis numa avenida de uma cidade para efeitos da correspondente regulação dos ciclos de semáforos nos cruzamentos com outras ruas (**Eng. Transportes**).

VII. Análise da distribuição espacial da qualidade de lodos de um rio, para vários níveis de profundidade, com base em amostras de sedimentos de distintas localizações, caracterizadas pelo seu grau de contaminação em diversos poluentes químicos e orgânicos (**Eng. de Georrecursos**). Tal análise por técnicas geoestatísticas reveste-se de importância para a definição de qual a estratégia de limpeza do leito do rio por dragagem em face do respetivo custo operacional.

O Controlo de Qualidade de produtos e serviços é um outro tópico crucialmente dependente da intervenção da Estatística, com enorme relevância em vários ramos da Engenharia e, em especial, da **Engenharia de Produção**.

A qualidade analisada estatisticamente é entendida no sentido de adequabilidade para uso, o que significa conformidade com as especificações requeridas. O Controlo Estatístico de Qualidade é então o conjunto de métodos estatísticos apropriado para medição, monitorização, controlo e melhoria da qualidade, tendo como objetivos fundamentais a redução da variabilidade dos resultados do processo, a eliminação de defeitos constatáveis e a otimização do seu desempenho. Tem como etapas importantes o Controlo Estatístico do processo em ação e o Planeamento Experimental exterior à produção visando a otimização ulterior do processo.

O Controlo Estatístico do Processo é assegurado por instrumentos estatísticos (diagramas) que permitam em tempo real detetar anomalias no processo, à medida que se vão analisando os resultados que vão saindo, e empreender as devidas ações corretivas para recolocar o processo sob controlo estatístico.

VIII. A título ilustrativo, suponha que uma empresa é contratada para produzir fios de cobre revestidos com um banho de prata para fins de aplicação em **Eletrónica**. O teor de prata nos fios é medido três vezes por dia durante um mês de trabalho. A análise ao longo do tempo das medições feitas através de diagramas apropriados permite constatar se o processo de aplicação do banho está ou não estatisticamente controlado.

Estatística em Ciências Sociais

As Ciências Sociais ao estudarem o comportamento humano em função do meio social com as suas relações de interdependência abrangem um largo leque de áreas como a Sociologia, Ciência Política, Psicologia, Demografia, Antropologia e Economia, entre outras, revelando ainda intercrossamentos com outros campos do conhecimento.

Não é por isso de estranhar que tal domínio possa fazer uso de metodologias múltiplas de análise que combinam técnicas qualitativas e quantitativas, em que nas últimas a Estatística tem sido chamada a intervir para dar resposta às questões solicitadas por vários estudos. Uma amostra destes é configurada pelos exemplos que se seguem:

I. Obter predições para as próximas eleições presidenciais a partir do registo histórico da percentagem de votos do candidato apoiado por um dado partido em cada um dos municípios do país, obtidos num conjunto de eleições passadas (**Ciência Política**).

II. Estudar o nível de cultura política dos cidadãos e sua relação com variáveis como o grau de instrução, tipo de emprego, nível económico, idade e género (**Sociologia**).

III. Caracterizar a privação, numa perspetiva multidimensional, dos agregados familiares de uma dada população, considerando múltiplos aspetos (capacidade económica, habitação, bens de conforto e formas de sociabilidade) determinantes do seu bem-estar e sua evolução ao longo de uma década (**Economia Social**).

IV. Analisar o desempenho de alunos em disciplinas fundamentais do ensino pré-universitário e o efeito nele do nível sócio-económico familiar e o tipo público ou privado da escola (**Ciências da Educação, Psicologia Social**).

V. Avaliar a capacidade de candidatos à admissão a um dado emprego através de respostas a questionários com múltiplos itens (**Psicometria**: ramo que estuda medições psicológicas – conhecimentos, competências, atitudes, crenças, inteligência - através de modelos e técnicas estatísticas e matemáticas).

Estatística em Finanças e Gestão Empresarial

I. Estudo do risco de mercado de um fundo de investimentos (**Econometria Financeira**)

Um dos objetivos em Finanças é a avaliação de risco de um ativo financeiro (ou carteira de ativos), risco este frequentemente medido em termos de variações de preços do ativo, como por exemplo os retornos.

A avaliação do risco de mercado de um ativo definido por um dado montante de milhares de euros em ações da empresa X, detido por um fundo financeiro Y, depende de dados sobre os preços diários passados dessas ações.

Supondo que se dispõe de uma série de retornos diários relacionados com os preços das ações da empresa, a sua análise estatística permite escolher uma gama de modelos com base nos quais se pode calcular e comparar os correspondentes valores em risco (o que se pode perder/ganhar) com uma dada probabilidade sobre horizontes temporais pré-especificados, e assim tomar decisões sobre a futura venda ou não ao longo desses períodos.

II. Estudo do campo da **Atuária**: análise de risco inerente à atividade seguradora e financeira (administração de seguros de vida/não vida e de fundos de pensões).

Análise estatística do montante total de indemnizações de cada tipo (danos corporais, danos materiais, danos próprios) -- função do número de sinistros e do montante envolvido em cada um destes -- pago por uma dada companhia de seguros do ramo automóvel durante uma sequência de anos. O objetivo central deste estudo é o da predição para o ano seguinte do montante total de indemnizações para efeitos de planeamento da política estratégica a seguir.

III. Análise de decisão e planeamento em meio empresarial (**Gestão de informação**)

O conhecimento do comportamento dos clientes de uma empresa é fundamental para a definição de estratégias de negócio. Por isso, as empresas apostam cada vez mais na construção de bases de dados complexas, contendo não só informação relativa às transações efectuadas por cada cliente, mas também informação complementar sobre este de forma a conhecê-lo cada vez melhor.

Usando a informação constante dessas bases de dados, o recurso a adequadas técnicas estatísticas permite, nomeadamente, segmentar clientes de acordo com os seus perfis de consumo de bens ou serviços e alocar novos clientes aos grupos criados, bem como identificar potenciais necessidades/apetências/desejos dos clientes, potenciando o desenvolvimento de novas vertentes de negócio.

Estatística em Direito

I. Estudos com implicações em **Genética Forense e Medicina Legal**.

Ainda que constituam domínios do conhecimento aparentemente afastados um do outro pelas matérias básicas que lhe são características, a Estatística e o Direito compartilham interesses fundamentais comuns ao lidarem com interpretação de evidência, testagem de hipóteses e tomada de decisões em contextos de incerteza. Tal facto, juntamente com a experiência acumulada no recurso a periciais avaliações estatísticas em processos judiciais, explica o uso crescente em algumas sociedades da Estatística na produção de provas de identificação para julgamentos.

Dados: Informação sobre testemunhos e vestígios encontrados em casos criminais (e.g., autoria de assassínios) e perfis de ADN em casos civis (testagem de paternidade em disputa).

Objetivos: Uso da informação adequadamente quantificada e de corretos raciocínios probabilísticos para avaliação da chance de identificação do(s) criminoso(s) (do verdadeiro pai) entre o(s) suspeito(s) (entre o(s) candidato(s) em disputa), de modo a reunir evidência suplementar e evitar argumentos enganadores tais como o da falácia do acusador (que consiste em interpretar incorretamente juízos probabilísticos condicionais por inversão do respetivo condicionamento). Isto coloca um enorme desafio ao trabalho em harmonia de estatísticos e advogados pelas suas distintas formações de base.

Estatística em Literatura e Linguística Quantitativa

I. Estudos de identificação autoral

No estudo quantitativo da Literatura e da Linguística Aplicada os métodos e modelos estatísticos são de importância hoje em dia inquestionável. A **Estilometria** (estudo quantitativo do estilo linguístico) ao lidar com a análise de textos para fins de identificação autoral de documentos anónimos ou disputados é um bom exemplo da aplicação da Estatística à área das Humanidades. A análise de textos pode ainda ser motivada para averiguação de plágio e confirmação ou eliminação de suspeitos de crime, sendo então integrável na denominada *Estilística Forense*.

Dados: Registo de frequências de determinadas palavras ou sequências de letras em várias parcelas dos textos a analisar, obtido pelo uso de computadores. O tipo de sequências mais usado pelo seu papel mais discriminador é o das palavras gramaticais (artigos, pronomes, preposições, conjunções) nas quais a função sintática é mais relevante que a função semântica, em oposição às palavras lexicais (substantivos, adjetivos, verbos, maioria dos advérbios). Outras variáveis possivelmente discriminativas para comparação de potenciais autorias ou de géneros literários (e.g., prosa versus poesia) são a extensão de palavras ou de frases e a proporção de pontuação (vírgulas, etc.).

Objetivos: caracterização do estilo literário de vários autores e atribuição de autoria em documentos por identificar, através de métodos estatísticos apropriados, hodiernamente aplicados através de software específico, em complemento a argumentos relacionados com as circunstâncias (pessoais, sociais, políticas, etc.).

Estatística no Desporto

A prática de virtualmente todas as provas desportivas acarreta a geração de números – de golos no futebol, de pontos no basquetebol, de tempos/distâncias em corridas/saltos de atletismo, pontos por júris em ginástica desportiva, etc.

A análise desses dados pode ser motivada por objetivos distintos que vão desde o simples entretenimento prazeroso de fãs de modalidades desportivas até à predição associada a interrogação sobre quais os limites de proezas atléticas humanas, passando pelo planeamento de melhores táticas de ação e estratégias de gestão, individuais ou coletivas, e pela necessidade de desmistificar conclusões empoladas pelos *media* (porque baseadas em pequenas amostras ou no manipular de níveis de desempenho extremos).

A Estatística tem efetivamente contribuído para

- I. a definição de adequada estratégia de gestão de uma equipa com base no estudo do efeito do montante relativo despendido com salários e transferências de jogadores na classificação final conseguida na liga por cada uma de várias equipas analisadas durante um conjunto de anos;
- II. a clarificação do efeito nos resultados das partidas de cada uma de várias equipas de expulsões ordenadas pelos árbitros e dos fatores de jogar como visitada versus visitante;
- III. o delineamento de estratégias ótimas a adotar por cada atleta de pista, tendo em conta o seu perfil fisiológico, e a atualização de sistemas de pontuação em modalidades como o decatlo, heptatlo, triatlo, e corrida de corta-mato por equipas;
- IV. evidenciar se as diferenças dos melhores desempenhos entre homens e mulheres nas provas de pista em atletismo têm vindo a esbater-se ao longo dos anos;
- V. prever a evolução de recordes por género com base em dados de recordes anteriores e de melhores marcas anuais e em modelos, com parâmetros de interpretação fisiológica e biomecânica, relacionando a distância e o tempo (ou velocidade) a percorrê-la.