Capitolo 1
La verifica della significatività dell'ipotesi nulla

1.1. Un ibrido non felice

Nell'analisi dei dati delle ricerche psicologiche il paradigma prevalente è la cosiddetta Verifica della Significatività dell'Ipotesi Nulla (VeSN — dall'inglese *NHST, Null Hypothesis Significance Testing*). La VeSN è il prodotto della fusione tra due approcci: uno di Sir R.A. Fisher (1925), il *p-value approach* (PVA); e uno di J. Neyman e E.S. Pearson (1933), the *fixed alpha approach* (FAA).

Di fatto, i due approcci hanno dei tratti in comune, al di là di differenze terminologiche e di quelle che sono, come vedremo, delle differenze sostanziali. Così entrambi condividono l'utilizzo dell'ipotesi nulla H_0 (peraltro non chiamata così da Fisher). Ciò che inoltre li accomuna è l'utilizzo di un valore critico di probabilità p (*p-value*) e di un livello critico αper determinare la probabilità del verificarsi di eventi dovuti al caso o ad errori di campionamento.

Di contro, gli stessi autori sopracitati non amavano particolarmente né l'altrui approccio, né l'ibrido risultante. Così, secondo Fisher (1955), il FAA è un approccio "russo" (inteso come "sovietico"), interessato soltanto all'efficienza, come nei piani economici quinquennali; e ognuno sa a quali risultati brillanti questi piani poi di fatto portavano. Ma secondo Jerzy Neyman (citato da Stegmüller, 1973), il PVA era "peggio che inutile".

È importante sottolineare che non sono molti i domini scientifici in cui viene applicata la VeSN, questo curioso prodotto dell'ingegnosità dell'uomo. È il caso ovviamente della psicologia, nonché della medicina (clinica e epidemiologica), della sociologia, dell'agricoltura, ma la maggior parte delle scienze (fisica, biologia, chimica, astronomia, e così via), peraltro, analizza i propri dati in modo alquanto differente. Esse preferiscono verificare dei *modelli* (vedi Cap. 6).

Nella maggior parte dei manuali di statistica applicata alla psicologia, la VeSN è usualmente presentata come *il* modo di analizzare i dati sperimentali. Si afferma che i test di significatività mirano ad ottenere informazioni riguardanti una certa caratteristica della popolazione che non è direttamente osservabile. Ciò che viene osservato sono le caratteristiche del campione, le quali permettono di fare inferenze sulla popolazione. Raramente viene detto esplicitamente che la verifica delle ipotesi è solo uno dei metodi che il ricercatore ha a disposizione per fare inferenze.

Bachmann Christina, Luccio Riccardo, Salvadori Emilia, *La verifica della significatività dell'ipotesi nulla in psicologia*, ISBN 88-8453-226-4,

Secondo l'approccio della VeSN (per una recente trattazione italiana del problema, vedi Pisati, 2002), il ricercatore formula due ipotesi mutuamente escludentisi: l'ipotesi nulla H_0 implica che il trattamento non abbia avuto effetto, mentre con l'ipotesi alternativa, detta abitualmente *sostantiva*, H_1, si ipotizza che ci sia stato un effetto. L'ipotesi nulla è chiamata anche ipotesi della non-relazione e non-differenza (Bakan, 1966; Cohen, 1988; Hinkle, Wiersma e Jurs, 1994); l'ipotesi sostantiva, detta anche *sperimentale*, indica che c'è una differenza tra le medie delle popolazioni dalle quali i campioni sono stati estratti, senza specificarne il verso (ipotesi bidirezionale), o specificandone la direzione (ipotesi monodirezionale). (Ciò vale, evidentemente, nel caso di disegni di ricerca che prevedano il confronto tra medie. Nel caso si abbiano disegni di tipo diverso — ad esempio correlazionali — l'ipotesi sarà differente, ma il senso del discorso non muta).

La logica che sottende la decisione statistica è di tipo falsificazionista: si parte dal presupposto che l'ipotesi nulla sia vera e si cerca di falsificarla. Solo se la probabilità *p* associata ai dati, ammesso che H_0 sia vera, è troppo bassa, si rifiuta H_0 in favore di H_1. Il limite entro il quale decidere se accettare o rifiutare H_0 viene fissato arbitrariamente, e si tratta del livello di probabilità α, detto anche livello di significatività critico, che rappresenta la probabilità teorica di rifiutare H_0 quando è vera. Se α viene fissato a 0,05, ciò significa che si ha la probabilità di rifiutare l'ipotesi nulla quando è vera in 5 casi su 100.

Calcolata la probabilità *p* associata ai dati osservati, ammesso che H_0 sia vera, con la statistica più appropriata, questa si confronta con α: se *p* è minore o uguale a α si rifiuta H_0, se *p* è maggiore di α si accetta H_0. Si può incorrere in due tipi di errore. Se si rifiuta H_0 quando è vera si commette un errore di I tipo (*falso positivo, falso allarme*), cioè si afferma erroneamente che il trattamento ha avuto effetto quando invece non lo ha avuto. Se invece si accetta H_0 quando è falsa si commette un errore di II tipo *(falso negativo,* detto anche *omissione*, con probabilità β), cioè si afferma che non c'è stato un effetto dovuto al trattamento mentre in realtà c'è stato.

Più è basso il livello di α e più è basso il rischio di commettere l'errore di I tipo, ma si badi che il livello di α non è indipendente da β, e cioè dalla probabilità di commettere un errore di II tipo; il rischio quindi è di sacrificare quella che è definita la *potenza* del test (e cioè, $1 - \beta$; vedi il Cap. 5). Ma, come vedremo meglio, ben pochi ricercatori si preoccupano di commettere questo tipo di errore, e di conseguenza nel tempo si è quasi smesso di prestare attenzione alla potenza statistica.

A chiarimento di quanto sopra, nei corsi di statistica per psicologi gli studenti vengono di solito posti di fronte, con piccole varianti, a questo schema (la cui efficacia didattica è peraltro indiscutibile).

Lo schema permette di visualizzare il processo decisionale. Si afferma che solo Dio sa se in natura sia vera H_0 o H_1. (Nell'Est europeo, se del caso, le eredità del *Diamat*, il materialismo dialettico, impongono di sostituire a Dio la Natura; cfr. Milosevic, 1995). Lo statistico allora gioca contro Dio (la Natura), e cerca di indovinare, sulla base della probabilità che assegna ai dati osservati, posto che H_0 sia vera, cosa c'è nella mente di Dio.

		Dio: H_0	Dio: H_1
Statistico	H_0	Rifiuto corretto	Falso negativo Errore di II tipo β
	H_1	Falso positivo Errore di I tipo α	Hit Potenza $(1 - \beta)$

Dato lo schema, allo studente viene allora detto che un processo inferenziale "genuinamente fisheriano" richiede i seguenti passaggi:

1. assegnare un valore di probabilità ad α (usualmente 0,05);
2. assumere che H_0 sia vera;
3. determinare la probabilità associata ai dati osservati ammesso che H_0 sia vera [cioè, p (dati | H_0)];
4. se $p > \alpha$, accettare H_0 (e il lavoro finirà con tutta probabilità nel cestino della carta straccia);
5. altrimenti, rifiutare H_0 e accettare H_1.

Lo studente non viene incoraggiato ad approfondire i problemi relativi ad errori di II tipo, potenza del test, e così via.

Ora, se il procedimento indicato può essere considerato effettivamente fisheriano, tutto l'apparato ha ben pochi rapporti con Fisher, ed è viceversa largamente dovuto a Neyman e Pearson (1933), i quali peraltro suggerivano una strategia decisionale sostanzialmente differente. È allora opportuno cercare di fare chiarezza, partendo dalla storia di questi straordinari personaggi che abbiamo evocato.

1.2. L'APPROCCIO FISHERIANO

1.2.1. Chi era Fisher

Partiamo proprio da Ronald Aylmer Fisher (1890-1962), non a torto definito frequentemente il "principe degli statistici" (un'ottima sua biografia è stata scritta dalla figlia Joan Box Fisher, 1978).

Fisher era nato da una famiglia benestante presso Londra, e fin da ragazzo dimostrò una straordinaria attitudine per la matematica. Peraltro, la morte della madre e la rovina economica del padre resero molto difficili gli anni della sua adolescenza. Egli studiò a Cambridge, al Gonville and Caius College, soprattutto matematica ed astronomia, ma si interessò anche alla biologia, e si appassionò (purtroppo) ai problemi dell'eugenica, coltivando delle detestabili idee politiche, ben entro i confini del razzismo.

Dopo la laurea, conseguita nel 1912, Fisher svolse diversi lavori da statistico e da insegnante di matematica nelle scuole secondarie, e cominciò, sia pure tra diverse incomprensioni, a farsi conoscere nel campo della statistica accademica, sinché, nel 1919, iniziò la sua carriera universitaria accettando la cattedra di statistica alla Rothamsted Agricultural Experimental Station, un importante istituto sperimentale di agraria, rifiutando nel contempo l'offerta di Karl Pearson, allora "Galton professor" all'University College di Londra (e incontrastato dominatore della statistica inglese, e non solo), di occupare l'importantissima posizione di statistico capo dei Laboratori Galton. (Per un vivace ritratto di Karl Pearson, vedi Williams, Zumbo, Ross e Zimmerman, 2003).

Si trattava chiaramente di una mano tesa da parte di Pearson, che se da un lato indicava una grande stima nei confronti del collega di tanto più giovane, dall'altro voleva essere un superamento delle polemiche che avevano diviso i due negli anni precedenti. Ma Fisher, dimostrando che l'asprezza del carattere sarebbe sempre stata un suo tratto distintivo, rifiutò l'offerta, portando a livello di totale rottura i rapporti con il collega più anziano.

La controversia tra i due, che ci interessa direttamente, dato che l'inimicizia di Karl Pearson per Fisher verrà ereditata dal figlio Egon, autore con Neyman dell'approccio alternativo al problema che qui trattiamo, e non sarà estranea allo sviluppo di questi due modi diversi di concepire l'analisi dei dati, nasce infatti nel 1917, quando Pearson criticò in un suo lavoro il concetto di "verosimiglianza" (*likelihood*), che era stato elaborato da Fisher per la prima volta nel 1915, e che dal 1921 in avanti sarebbe stato da quest'ultimo considerato uno dei suoi più importanti contributi alla statistica. Fisher respinse aspramente le critiche di Pearson, e la sua irritazione si accrebbe quando seppe che nel 1918 Pearson era stato uno dei *referees* che avevano fatto rifiutare un suo lavoro inviato alla Royal Society (anche se di fatto Pearson, trattandosi di problemi di statistica applicata alla genetica, aveva dato parere favorevole per ciò che riguardava gli aspetti statistici, mentre aveva chiesto un giudizio da parte di persona più esperta per quel che riguardava gli aspetti biologici).

Ma il rifiuto del 1919 creò una situazione di guerra aperta. Mentre Fisher, alla stazione Rothamsted metteva a punto l'analisi della varianza (ANOVA — il suo contributo più noto, se non il più importante, alla statistica del XX secolo), e pubblicava altri studi di fondamentale importanza, tra cui merita di essere segnalato quello sul concetto di informazione, del 1925, Pearson iniziava la sua offensiva aperta contro di lui. In un articolo del 1922 lo accusò di avere usato il χ^2 (una statistica da Pearson stesso creata) in modo erroneo, e approfittò della sua posizione di presidente della Royal Statistical Society e di *editor* di *Biometrika*, la più importante rivista di statistica di allora, per bloccare la pubblicazione di molti articoli di Fisher, che nel 1925 si dimise per protesta dalla Società. Ma Fisher, evidentemente, non rimase con la "penna in

mano", e la polemica tra i due crebbe sino a livelli insostenibili, ed ebbe termine solo con la morte di Karl Pearson, nel 1936.

C'è un'ironia della sorte nel fatto che quando Pearson, nel 1933, andò in pensione, la sua cattedra venisse assegnata proprio a Fisher. Ma poiché il destino deve sempre creare situazioni complesse, e possibilmente sgradevoli, la cattedra fu divisa a metà: la parte di Eugenica fu per Fisher, ma quella di statistica andò al figlio di Karl, Egon Sharpe Pearson, che, come vedremo meglio nel prossimo paragrafo, si era già segnalato per la sua ostilità, largamente ricambiata, nei confronti di Fisher. Di più, anche il Dipartimento venne diviso, e Pearson fu fatto Direttore del nuovo dipartimento di Statistica Applicata.

Gli anni '30 furono quindi anni di conflitto, esacerbato anche dalle inaccettabili concezioni politiche generali che Fisher man mano sviluppava. In particolare, suscitò violente polemiche il suo darwinismo sociale, in base a cui sosteneva ad esempio l'opportunità di sterilizzare i deboli mentali, e la necessità che la società rendesse difficile l'esistenza, con un sistema premiante in termini di reddito e benefici, agli individui sterili e meno sani. Queste idee non erano peraltro nuove per Fisher. Già nel 1924, in uno dei suoi meno degni lavori, aveva sostenuto la necessità di sterilizzare i deboli mentali, tentando di dimostrare la falsità del cosiddetto "principio di Handy-Weinberg" (dal nome dei due studiosi che lo avevano scoperto indipendentemente nel 1908), secondo cui è impossibile con la sterilizzazione eliminare da una popolazione i geni recessivi. È peraltro corretto dire che si trattava di idee che se oggi ci appaiono inaccettabili, all'epoca erano largamente diffuse, almeno tra i genetisti. Peraltro Fisher non ebbe mai paura di esprimere idee politicamente indecenti, fino al suo tentativo di screditare le basi statistiche delle ricerche che cominciavano a dimostrare la relazione causale tra fumo e cancro ai polmoni (1958).

Nel 1943, Fisher decise infine di lasciare Londra per Cambridge, dove fu chiamato come "Balfour Professor" di Genetica, e dove rimase fino al 1959, due anni dopo la pensione, seguitando ad insegnare in attesa che venisse nominato un suo successore. Fu tra l'altro qui nominato Presidente del Gonville and Caius College, dove aveva studiato. Si trasferì poi all'Università di Adelaide, in Australia, dove morì nel 1962.

Tra le numerosissime opere di Fisher, quelle che forse hanno avuto il maggior impatto nel mondo della ricerca sono state *Statistical Methods for Research Workers* (1925), che ebbe un numero infinito di riedizioni, e che ha radicalmente indirizzato in modo nuovo, sino a tutt'oggi, l'analisi dei dati non solo in biologia e in agricoltura, ma anche in psicologia e in sociologia; e *The Design of Experiments* (1935), che è stato la guida della maggior parte degli statistici nell'applicazione dell'ANOVA.

1.2.2. Il p-value approach

L'enunciazione più chiara del *p-value approach* di Fisher si trova proprio nel più volte citato *Statistical Methods for Research Workers.* Qui, nel primo capitolo introduttivo (p. 9), Fisher afferma che un problema della statistica, "su cui fino ad ora sono stati compiuti pochi progressi, è alla base dei test di significatività, con i quali noi possiamo

esaminare se i dati sono o meno in armonia con le ipotesi che vengono avanzate". Se noi traiamo un campione da un universo, le caratteristiche del campione possono farci prevedere quali sono le caratteristiche dell'universo; se quindi un secondo campione viola le nostre aspettative, "possiamo inferire che è stato tratto da una diversa popolazione […] Test critici di questo tipo possono essere chiamati test di significatività, e se tali test sono disponibili possiamo scoprire se il secondo campione è o meno significativamente diverso dal primo" [p. 44].

Poco oltre, compare questo limite di probabilità di 0,05, che trionferà poi nei decenni successivi come signore incontrastato dell'inferenza statistica. Dalle tabelle della curva soggiacente alla distribuzione normale, Fisher mostra che a un valore di unità di deviazione standard di 1,96 corrisponde a 2 code una probabilità di 0,05 (o di uno su venti). "È conveniente assumere questo punto come limite per giudicare se una deviazione va considerata o meno significativa. Le deviazioni che eccedono il doppio della deviazione standard sono così considerate formalmente significative" [p. 48]. Si osservi il contrasto tra il "conveniente" e il "formale".

Nel Cap. 4, dedicato al χ^2, al § 21 viene presentato il problema della verifica dell'indipendenza tra variabili nelle tabelle di contingenza, in termini in cui la sola differenza che possiamo trovare rispetto ai trattamenti di oggi è probabilmente data dalla straordinaria chiarezza e sequenzialità delle argomentazioni di Fisher (che per una curiosa leggenda metropolitana seguita a essere definito "di difficile lettura", se non "oscuro").

Nel precedente § 20, Fisher ha mostrato l'uso del χ^2 per verificare la bontà dell'adattamento di dati osservativi a una funzione. In questo paragrafo, affronta "una classe particolare e importante di casi in cui l'accordo tra aspettativa ed attesa che può essere sottoposta a verifica comprende dei test di *indipendenza*. Se lo stesso gruppo di individui viene classificato in due (o più) modi diversi, per esempio persone inoculate e non inoculate, nonché attaccate e non attaccate da una malattia, può essere necessario sapere se le due classificazioni sono indipendenti" (p. 81, corsivo di Fisher). Fisher presenta poi degli esempi numerici (che oggi eviteremmo con cura: l'N complessivo del suo primo esempio è di 18.483 individui, e ben sappiamo quanto il χ^2 sia sensibile alla grandezza del campione!), in cui mostra come calcolare le frequenze attese e come confrontarle con quelle osservate, ai fini del calcolo della statistica, nonché le formule abbreviate. Fisher giunge così a un esempio in cui esamina una tabella 2×2, e afferma: "I valori attesi sono calcolati dal totale osservato, e le quattro classi [le quattro frequenze di cella; NdR] devono avere una somma concordante, e se i valori di tre classi sono date in modo arbitrario il quarto valore è perciò determinato, di qui n=3, χ^2=10,87, e la probabilità di eccedere tale valore è compresa tra 0,01 e 0,02*; se assumiamo* p = *0,05 come limite della deviazione significativa, diremo che in questo caso le deviazioni dall'aspettativa sono chiaramente significative*." [pp. 82-83, corsivo nostro].

Nell'accennare in altre parti dei *Methods* al problema della significatività, Fisher non si sposta da questa posizione. La probabilità va posta a 0,05 perché è conveniente farlo, ed è un livello sufficiente per prendere una decisione ragionevole.

Si osservi, peraltro, che nel corso della sua attività scientifica Fisher ebbe modo di cambiare opinione almeno su un punto non secondario del suo approccio. Infatti, se nel 1925 appare chiaro che la probabilità va posta *a priori* a 0,05, negli anni '50

(Fisher, 1955, 1956) egli afferma esplicitamente che l'esatto livello di significatività deve essere calcolato *dopo* che si è effettuato il test. Il livello di significatività è quindi una proprietà dei dati. Anche in questo senso il *p-value approach* si differenzia dall'approccio di Neyman e Pearson, che vedremo ora.

1.3. Il *fixed-alpha approach*

1.3.1. Egon Pearson e la sua controversia con Fisher

Come abbiamo avuto modo di accennare, la controversia tra Egon Sharpe Pearson e Ronald Aylmer Fisher esplose ben prima che i due entrassero in concorrenza diretta per la successione alla cattedra di Karl Pearson. Vi fu da un lato il desiderio del primo di difendere il padre dai furibondi attacchi che Fisher seguitava a rivolgergli, dall'altro l'insofferenza del secondo per qualsivoglia critica gli venisse rivolta. Lo scontro iniziò negli anni '20, con due recensioni che Pearson (1927, 1929) scrisse alle prime due edizioni del volume di Fisher (1925) sugli *Statistical Methods for Research Workers.*

L'immagine che ci viene solitamente tramandata di Egon Pearson è quella di un personaggio abbastanza grigio, schiacciato da un lato affettivamente da un padre straripante, che lo iperprotesse e gli prefigurò una carriera straordinaria, con pochi meriti da parte del giovane Egon; e dall'altro schiacciato intellettualmente da quell'altra grande figura che fu Jerzy Neyman, certamente uno dei più grandi matematici applicati del '900, a cui si devono in larghissima misura le basi teoriche dell'approccio che va sotto i loro due nomi. In realtà, Egon Pearson fu una personalità abbastanza forte, certo non prepotente quanto il padre, ma perfettamente in grado di difendersi ed attaccare ove necessario. E certamente i suoi contributi originali alla statistica, seppure non geniali quanto quelli di Neyman, furono tutt'altro che disprezzabili. Ma di sicuro Pearson fu un grande organizzatore del lavoro scientifico, e forse ancor più di Fisher contribuì a collocare la scuola statistica britannica in quella posizione di eccellenza che ancor oggi occupa.

È comunque indubbio che il padre lo agevolò non poco. Iscrittosi all'Università, al Trinity College di Cambridge, nel 1914, evitò il servizio militare per un soffio al cuore (era appena scoppiata la I Guerra Mondiale), ma volle rendersi utile al Paese servendo come civile all'Ammiragliato, per cui interruppe gli studi. Riuscì però a conseguire la laurea nel 1919 in una speciale sessione per militari, a cui non avrebbe avuto molto diritto di partecipare, ed entrò come ricercatore in astrofisica a Cambridge, cominciando ad interessarsi alla teoria degli errori.

Tanto bastò perché nel 1921 Karl Pearson riuscisse a fare entrare il figlio come *lecturer* nel suo Dipartimento di Statistica Applicata all'University College di Londra. In realtà, la storia dice che di suo in quel periodo di lezioni non ne fece, facendo invece le lezioni al posto del padre nel corso di questi. E così il padre lo fece diventare nel 1924 vice direttore di *Biometrika*, la rivista più importante di statistica dell'epoca, che dirigeva, e che era nelle mani di Karl Pearson uno straordinario strumento di potere.

La svolta ci fu nel 1925, quando Jerzy Neyman giunse all'University College grazie a una borsa Rockfeller biennale. Di questo arrivo di Neyman parleremo più distesamente in seguito, ma è certo che tra i due nacque una profonda amicizia, cementata da una grande affinità intellettuale. Soprattutto, però, il lavoro con Neyman diede una nuova sicurezza al giovane Pearson, anche se di certo si guardò bene dal tagliare il cordone ombelicale con il padre.

Nel 1926 Pearson pensò bene di entrare direttamente nell'agone contro Fisher, e lo fece con una perfida, brevissima recensione al saggio sugli *Statistical Methods* del 1925 di questi, che appena pubblicato ebbe da subito una risonanza enorme. La recensione di Pearson è apparentemente rispettosa. Il compito che Fisher si è assunto è "molto difficile", la sua stesura del testo è "diligente", ma è "un po' dubbio" che il lettore possa poi applicare i suoi risultati a situazioni "un po' diverse" dagli esempi che porta. Soprattutto, poi, è difficile "seguire le sue dimostrazioni basate sul concetto di gradi di libertà […] che appaiono fondarsi su analogie". Di più, un metodo "consolidato da tanto tempo" come quello del rapporto di correlazione, è incomprensibilmente "accennato di sfuggita". Ma "se i vecchi metodi sono trattati solo sommariamente", esempi e nuove tabelle sono di "considerevole interesse".

Ce n'era a sufficienza per fare infuriare Fisher, specie se si tiene conto del fatto che la sua polemica sempre più aspra con Pearson padre aveva avuto al suo centro proprio il problema dei gradi di libertà, specie nel χ^2, e della correlazione. La risposta (Fisher, 1927) non si fece attendere. Manca qui qualsiasi cortesia formale. Fisher non si aspetta che Pearson "sia d'accordo", ma fa presente che già nel 1922 aveva dimostrato tutti i limiti del coefficiente di correlazione, e le tre pagine dedicategli nel libro (errore semmai di "commissione e non di omissione") servivano a evitare agli studiosi perdite di tempo e a cadere nelle incongruenze di cui erano rimasti vittime tanti "illustri biometristi"; e l'allusione a Karl Pearson non poteva essere più trasparente.

Pearson attese l'uscita della seconda edizione (1928) degli *Statistical Methods* per tornare all'attacco, e questa volta lo fece dalle colonne di *Nature*, ben più influenti di quelle di *Science Progress*, su cui era comparsa la prima polemica (Pearson, 1929a). Qui la recensione non era firmata, ma l'autore era riconoscibilissimo. Il problema al centro dell'interesse di Pearson era quello dell'applicabilità di tanti test statistici concepiti per variabili distribuite normalmente quando venivano usati su campioni di cui non era nota l'appartenenza a distribuzioni normali, o che comunque si dipartivano dalla normalità. È questo il problema della robustezza delle statistiche, a cui Pearson si stava comunque appassionando, e sui cui stava stimolando l'interesse dell'amico Jerzy Neyman, rientrato in Polonia, un problema tutt'oggi di grande attualità, ad esempio nell'analisi delle strutture di covarianza.

Questa volta Fisher divenne letteralmente furioso. Invano tentò di far da paciere William Sealy Gossett, questa curiosa figura di "birraio" (diresse per alcuni decenni la famosa Birreria Guinness), chimico con la passione per la statistica, famoso ancora adesso con lo pseudonimo di "Student" — il test *t* di Student è forse, con il χ^2 la statistica più usata nella ricerca inferenziale. Student aveva buoni rapporti con entrambe le parti in causa (Pearson, 1990, gli avrebbe dedicato una bellissima biografia); vi fu, tra lui e Fisher, un frenetico scambio di corrispondenza, e infine Gossett pensò di

chiudere l'incidente con una lettera a *Nature* (Gossett, 1929), in cui, in modo un po' contorto, sosteneva che (i) il recensore (Pearson) aveva sollevato il fatto che un lettore disattento avrebbe potuto credere che Fisher sostenesse che il suo lavoro si poteva applicare tranquillamente anche a distribuzioni non normali; (ii) che il recensore, usando il verbo "ammettere" (*admit*), anziché ad esempio "sottolineare" (*stress*), fa credere che Fisher sia caduto nell'equivoco; (iii) che comunque egli riteneva che ad esempio la distribuzione di Student potesse essere usata ragionevolmente bene anche per dati non distribuiti normalmente; e infine, (iv) che era comunque opportuno che Fisher stesso potesse indicare, dandone le basi teoriche, quali modifiche dovessero essere fatte alle sue tabelle nel caso di scostamenti dalla normalità.

Ma Fisher non si quietò affatto, e in una sua lettera a *Nature* (Fisher, 1929) se la prese equamente non solo con Pearson, ma anche con Student. In che senso avrebbe mai dovuto dire come correggere le sue statistiche e le sue tabelle per condizioni di non normalità? Non ci si rendeva conto che non ci sarebbe mai stato un limite alle correzioni da apportare? Non si trattava tanto di "stoltificazione" di qualsiasi metodo statistico, ma di "abbandono della teoria degli errori". E, in ogni caso, nella pratica della ricerca erano problemi che non si ponevano.

La palla tornava allora a Pearson, che (1929b) ribadiva, con un esempio tratto da un articolo su *Biometrika* sulla lunghezza di uova di alcune specie di uccelli, che il caso in cui si avevano scostamenti anche sensibili dalla media anche in biologia era tutt'altro che eccezionale, e i metodi di Fisher non potevano darne conto.

Si osservi che in tutto questo l'atteggiamento di Egon Pearson era ben diverso da quello del padre Karl. Come nota Reid (1997), temperamento timido e introverso, sempre tormentato da un acuto sentimento di inferiorità, specie nei confronti del padre, nel corso della controversia, come negli anni a venire, Egon soffriva del tormento di scoprire che il padre poteva avere torto; ma date anche le sue modeste basi matematiche, non tutto capiva di quello che Fisher scriveva, e comunque lo odiava per i continui attacchi al padre, che non cessarono neppure dopo la morte di questi. Eppure sentiva che Fisher spesso aveva ragione.

Si può facilmente immaginare quanto poco sereno dovesse essere il rapporto tra Egon Pearson e Fisher quando nel 1933, come si è avuto modo di accennare, al ritiro di Karl Pearson la sua cattedra venne divisa tra i due. Peraltro, grazie soprattutto alla collaborazione con Neyman, Egon Pearson non era più il figlio di papà che aveva fatto carriera solo grazie agli appoggi paterni, e poteva vantare una serie di contributi scientifici di tutto valore.

Gli anni che vanno dal 1926 al 1933 sono i più produttivi della sua attività scientifica. Sono gli anni della collaborazione con "Student", che tentò invano di rendere più sereni i rapporti con Fisher, ma che certamente insegnò a Egon infinitamente di più di quanto era riuscito a insegnargli il padre. Ma soprattutto sono gli anni dell'intensa corrispondenza con Jerzy Neyman, che si coronerà nel 1933 con il famoso lavoro sul "lemma" (vedi oltre). Neyman tornò in Inghilterra nel 1934, per poi lasciarla definitivamente per Berkeley nel 1938. Curiosamente, al suo ritorno la collaborazione tra i due si affievolì. Tra l'altro, alla sua morte Neyman (1981) ha fatto presente che negli anni della maggiore collaborazione l'iniziativa era quasi sempre stata di Pearson.

Questi si interessò ora soprattutto ad aggiornare le opere del padre, specie le famose *Tables for Statisticians and Biometricians.* Uno sguardo alle sue opere più significative, raccolte nel 1966, ci mostra che, anche se non mancarono dei contributi pregevoli, il suo periodo di grazia si era comunque concluso. Ne rimase l'immagine di un gentiluomo mite e affabile con tutti e da tutti apprezzato (salvo, naturalmente, che dall'infame Fisher), grande organizzatore del lavoro scientifico, propalatore instancabile dei metodi statistici applicati alla ricerca empirica.

1.3.2. Jerzy Neyman

L'altro grande personaggio all'interno della controversia, e a giudizio di chi scrive forse il più grande e sotto ogni aspetto il più rispettabile dei tre, era un matematico moldavo, Jerzy Splawa Neyman (1894-1981; il primo cognome, Splawa, di origine nobiliare, fu da lui abbandonato negli anni '20).

Neyman nacque a Bendary, allora appartenente alla Russia, da una famiglia cattolica polacca, e studiò a Kharkov, dove la madre in difficoltà economiche si era trasferita nel 1906 dopo la morte del marito. All'Università studiò dal 1912 al 1917 fisica e matematica, appassionandosi soprattutto alla teoria della misura di Lebesgue, su cui, ancora studente, scrisse i primi saggi scientifici. Ma la lettura che dovette segnare la sua vita fu *Grammar of Science* di Karl Pearson, segnalatagli da A. Bernstein, suo professore di Probabilità, un'opera che per il suo antidogmatismo lo doveva prima turbare profondamente, poi affascinare. Tra l'altro, questo libro, di immensa diffusione, caratterizzato anche da un'impostazione materialistica, ricevette dei giudizi estremamente lusinghieri da parte di Lenin.

Dopo la laurea rimase all'Università di Kharkov come assistente, e iniziò ad appassionarsi di statistica; ma lo scoppio della guerra e la Rivoluzione russa resero le condizioni di vita particolarmente dure. Lo scoppio della guerra russo-polacca fece quindi precipitare la situazione. Considerato polacco dai russi, fu prima arrestato e quindi preferì trasferirsi con la moglie, sposata due anni prima, nel 1921 in Polonia. Qui con l'aiuto di Sierpinski, dopo qualche lavoro saltuario come statistico, nel 1923 riuscì a entrare all'Università di Varsavia come assistente. Nel 1925, poi, come abbiamo già avuto modo di dire, grazie a una borsa biennale Rockfeller potè raggiungere Londra, e soprattutto poté andare a lavorare con quel Karl Pearson che tanto lo aveva intellettualmente affascinato.

L'impatto, peraltro, non fu dei più felici. Da molti anni Karl Pearson aveva smesso di aggiornarsi scientificamente, troppo preso dalla gestione del suo immenso potere, e Neyman rimase sconvolto dall'ignoranza dei più recenti risultati in campo matematico che regnava all'University College. L'aspetto positivo fu però dato dall'amicizia che poté stabilire con Egon Pearson, premessa di una straordinaria collaborazione tra i due.

In ogni caso, Neyman ritenne più opportuno utilizzare la sua borsa per andare a Parigi, e lì poté incontrare, a un livello di gratificazione molto più elevato, un altro idolo dei suoi anni universitari, Lebesgue, oltre ad altri grandi matematici dell'epoca,

come Borel e Hadamard. Lì fu raggiunto da una lettera di Egon Pearson, che lo invitava a collaborare con lui su problemi di statistica. Pearson lo raggiunse anche a Parigi nel 1927 per un breve periodo, e ancora in Polonia, dove Neyman tornò nel 1928, per brevi scambi diretti di idee, oltre all'intensa corrispondenza. La collaborazione si rivelò straordinariamente proficua. Neyman poteva offrire a Pearson quelle basi matematiche in cui il giovane inglese si mostrava carente, e questi a sua volta viveva in un osservatorio privilegiato, avendo quotidianamente sotto gli occhi quelli che erano i problemi più vivi del dibattito statistico del tempo. In particolare, non era tanto l'influenza del padre (le cui carenze Egon cominciava a scoprire), quanto quella di "Student", e soprattutto quella del nemico Fisher, che Pearson studiava accanitamente tra infinite difficoltà matematiche, che stimolavano i due a quel loro straordinario impegno.

Nel 1933 i due lavori sul "lemma" erano una realtà. Nello stesso anno, diventato direttore del Dipartimento di Statistica Applicata dell'University College, Pearson poteva invitare Neyman a Londra, dove giunse nel 1934, e che abbandonò nel 1938 per l'Università di California a Berkeley. Negli anni di Londra Neyman fornì dei contributi fondamentali sul campionamento e sugli intervalli di fiducia. E soprattutto un lavoro di straordinaria importanza, nel 1937, sulla stima statistica.

Neyman non lasciò più Berkeley, e vi fondò nel 1955 un Dipartimento di statistica. Peraltro, lasciata Londra, i contributi più importanti di Neyman furono destinati alle applicazioni della statistica, dalle elezioni alla medicina alla meteorologia.

1.3.3. Il lemma di Neyman e Pearson

Di fatto, lo schema presentato sopra non è fisheriano, ma deriva dalla FAA di Neyman-Pearson. I concetti di ipotesi alternativa, errore di I e di II tipo, e di potenza statistica non sono frutto della teorizzazione fisheriana, ma sono stati formulati da Neyman & Pearson (1933). Al contrario di Fisher, questi autori vedevano i test di significatività come un metodo per selezionare una ipotesi tra due ipotesi possibili e non come procedura di *testing* di una sola ipotesi. Si osservi che, come già rilevato, nell'ottica fisheriana l'ottenere un p associato ai dati, ammesso che H_0 sia vera, e cioè p(dati H_0), inferiore al livello di α prefissato, porta a rifiutare l'ipotesi nulla, poiché un valore di probabilità così basso rende implausibile la condizione della sua verità. Ma p non fornisce nessuna indicazione sulla verità di H_1.

Nella FAA, di contro, in primo luogo è fissato *a priori* il valore di α (di qui il nome dell'approccio). Infatti Neyman e Pearson ritengono che il primo problema è quello di evitare gli errori di I tipo, ma prestano altrettanta attenzione agli errori di II tipo. Il loro procedimento prevede che la decisione che viene presa, che non è del tipo "tutto o nulla", sia quella che preserva la massima potenza possibile del test, $1 - \beta$ una volta fissato α. Va peraltro osservato che questa diversa impostazione non riguarda solo aspetti di natura statistico-matematica, ma trova le sue radici in un modo profondamente diverso di concepire il ruolo che questi studiosi attribuivano alla loro disciplina, alla ricerca scientifica, e in generale alla visione del mondo. E in parte certe differenze derivavano anche da problemi più personali.

Vediamo allora di presentare l'approccio di Neyman e Pearson in quella che è la sua forma più nota, e cioè il loro famoso "lemma". Di questo lemma daremo una presentazione per quanto possibile poco tecnica, anche se cercheremo di mantenerci a livello di rigore.

1.3.3.1. La ripartizione dello spazio dei parametri

Il punto di partenza è quello della definizione di una popolazione data da una variabile casuale continua ξ, che si distribuisce sulla base di una funzione di densità di probabilità $z(\xi; \theta)$. I valori che questa variabile assume li indicheremo con x. Si osservi che θ è il parametro, o i parametri, della funzione z e appartiene allo spazio dei parametri Ω, e cioè $\theta \in \Omega$.

Vediamo questo cosa significa, nel caso di due tra le funzioni di densità di probabilità di riscontro più frequente in psicologia, e cioè la funzione normale e la funzione di Poisson. Nel primo caso, vi sono due parametri, che sono rispettivamente la media μ e la deviazione standard σ, e quindi abbiamo:

$$z(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]}, \tag{1.1}$$

e quindi la funzione ha due parametri, media μ e deviazione standard σ.

Nel caso della distribuzione di Poisson, abbiamo invece un solo parametro, λ, essendo la funzione

$$z(x;\lambda) = e^{-\lambda} \frac{\lambda^x}{x!}. \tag{1.2}$$

(Con l'aumentare di λ la distribuzione diventa sempre più simmetrica).

Il passo successivo consiste nella possibilità di operare una partizione nello spazio dei parametri Ω. Immaginiamo che tale partizione ci consenta di dividere questo spazio in un sottospazio C e nel sottospazio complementare $A = \Omega - C$. A può essere a sua volta ripartito in altri sottospazi, e chiameremo uno qualsiasi di questi sottospazi C'. Chiamiamo i due sottospazi C e A, perché il primo per motivi che saranno più chiari in seguito rappresenta la regione critica, il secondo la regione di accettazione dell'ipotesi nulla.

Ora, in cosa consiste la decisione statistica? Noi, è noto, non abbiamo a disposizione la popolazione, ma solo dei campioni. Facciamo allora il caso più semplice: abbiamo estratto un campione, che rappresenteremo come un vettore $\mathbf{x} = [X_1, X_2, \ldots, X_n]$, per cui dovrà valere la nostra funzione di densità di probabilità $z(\mathbf{x}; \theta)$. Nota bene: noi ci poniamo il problema se questo campione appartiene a questa popolazione; per esempio, se questo campione, che ha assunto un sonnifero, dorma in modo uguale

o diverso dalla quota della popolazione che il sonnifero non lo ha assunto. Per fare questo, partiamo dall'assunto che la legge z valga allo stesso modo e per il campione e per la popolazione, altrimenti il confronto sarebbe insensato. Ma se la legge è la stessa, quello che cambia è il parametro (i parametri). Nell'esempio fatto, il problema che ci poniamo è se chi assume il sonnifero dorme di più di chi non lo assume. Posto che popolazione e campione si distribuiscano entrambi normalmente, andremo allora a vedere il parametro media μ nella popolazione e nel campione.

In altri termini, il valore del parametro sarà lo stesso in tutti i sottospazi di Ω in assenza di trattamento, o nel caso che per ipotesi il trattamento sia inefficace. Quest'ultima ipotesi è quella che abbiamo chiamato ipotesi nulla H_0 nel primo paragrafo del capitolo. Possiamo formalizzare quanto sopra in questi termini: viene stabilita una funzione di ripartizione di Ω $t(\mathbf{x})$ tale per cui, posto che H_0 sia vero, la probabilità condizionale che θ appartenga a C è uguale alla probabilità condizionale che appartenga a qualsivoglia altro sottospazio C' di Ω, e questa probabilità è appunto l'α prefissato; e cioè:

$$p\left[t(\mathbf{x}) \in C \middle| \mathrm{H}_0\right] = p\left[t(\mathbf{x}) \in C' \middle| \mathrm{H}_0\right] = \alpha. \tag{1.3}$$

Ciò posto, nel caso che sia invece vera l'ipotesi sostantiva H_1, il nostro parametro verrà ad avere un valore diverso nel sottospazio C rispetto a quello che ha negli altri sottospazi, e l'applicazione della funzione di ripartizione deve portare a un valore di probabilità condizionale più alto quando siamo in C rispetto a quando siamo in un qualsiasi altro sottospazio C':

$$p\left[t(\mathbf{x}) \in C \middle| \mathrm{H}_1\right] > p\left[t(\mathbf{x}) \in C' \middle| \mathrm{H}_1\right]. \tag{1.4}$$

1.3.3.2. La stima dei parametri

Prima di passare al lemma propriamente detto, il nostro problema è però quello di introdurre rapidamente i concetti di stima dei parametri e di massima verosimiglianza. Evidentemente non potremo trattare che molto superficialmente un argomento così complesso, ma quel che ci interessa è chiarire qualche concetto, tra cui quello di massima verosimiglianza, sviluppato da Fisher (1921a), ed estremamente importante per il lemma di Neyman-Pearson.

La *funzione di verosimiglianza* può essere definita come la funzione di densità di probabilità congiunta di n variabili estratte casualmente (e quindi indipendenti) dall'universo (e quindi identicamente distribuite), definita nello spazio Ω. Se la funzione di densità di probabilità è $z(\mathbf{x};\ \theta)$, la funzione di verosimiglianza, funzione di densità di probabilità congiunta, sarà quindi, date le n variabili estratte $(x_1, x_2, \ldots, x_n)$,

$$L(x;\theta) = z(x_1;\theta)z(x_2;\theta)...z(x_n;\theta) = \prod_{i=1}^{n} z(x_i;\theta). \quad (1.5)$$

Si osservi che di solito si utilizza, per motivi soprattutto di semplicità di calcolo, non la funzione L di verosimiglianza, ma la sua trasformazione logaritmica G, e cioè:

$$G(x;\theta) = \ln L(x;\theta) = \sum_{i=1}^{n} \ln z(x_i;\theta). \quad (1.6)$$

Tutto ciò cosa significa? Il nostro compito è quello di trovare la migliore stima per il parametro che ci interessa. In questo caso, si badi, θ è la variabile, mentre i diversi valori di x estratti possono essere considerati costanti. La migliore stima del parametro è allora quella per cui è massima la probabilità che i valori che costituiscono il vettore della variabile casuale campionaria $\mathbf{x}$ assumano i valori particolari $(x_1, x_2, \ldots, x_n)$. Si tratta allora di trovare il valore di θ tale per cui $G(\mathbf{x}; \theta)$ è massimo. Tale valore, indicato come $\hat{\theta}$, è detto "stima di massima verosimiglianza", ed è diverso da qualsiasi altro possibile valore del parametro $\breve{\theta}$. In altri termini,

$$G(\mathbf{x};\hat{\theta}) \geq G(\mathbf{x};\breve{\theta}), \quad (1.7)$$

per qualsivoglia $\breve{\theta}$ diverso da $\hat{\theta}$.

Per trovare allora $\hat{\theta}$ dobbiamo ricorrere all'analisi. Noi sappiamo che il punto di massimo di una funzione è quello in cui si annulla la sua derivata prima, e in cui la sua derivata seconda è negativa. In altri termini, ci basterà trovare il valore di θ tale per cui

$$G'(\mathbf{x};\theta) = 0, \quad (1.8)$$

$$G''(\mathbf{x};\theta) < 0. \quad (1.9)$$

Possiamo così ora affrontare il lemma di Neyman-Pearson.

1.3.3.3. Il lemma

Veniamo quindi al lemma di Neyman-Pearson propriamente detto. Estraiamo, come si è detto, dalla popolazione ξ che ha funzione di densità di probabilità $z(\xi, \theta)$, un campione $\mathbf{x} = [X_1, X_2, \ldots, X_n]$, per cui varrà comunque la stessa funzione di densità di probabilità, $z(\mathbf{x}; \theta)$. Formuliamo quindi le nostre ipotesi, nulla e sostantiva, in forma un po' diversa da quella a cui siamo abituati, e cioè

$$H_0 : \theta = \theta_0 \tag{1.10}$$

$$H_1 : \theta \neq \theta_1 \tag{1.11}$$

In realtà, ogni impressione di stranezza scompare, solo che poniamo mente al fatto che θ potrebbe essere μ. In altri termini, l'ipotesi nulla afferma che la migliore stima del parametro corrisponde al valore del parametro dell'universo, mentre l'ipotesi sostantiva afferma che ne differisce.

Come stabilirlo? Detta G (o L, il che ai nostri fini è lo stesso) la funzione di verosimiglianza, assumiamo di poter stabilire un valore costante k_α, tale per cui la probabilità di errore di primo tipo sia uguale ad α. In altri termini, definite le funzioni di verosimiglianza per l'ipotesi nulla e l'ipotesi sostantiva, rispettivamente $G(\mathbf{x}; \theta_0)$ e $G(\mathbf{x}; \theta_1)$, α deve essere uguale alla probabilità condizionale che il loro rapporto sia minore di k_α, posto che H_0 sia vero:

$$p\left[\frac{L(\mathbf{x};\theta_0)}{L(\mathbf{x};\theta_1)} < k_\alpha \,|\, H_0\right] = \alpha. \tag{1.12}$$

Allora la regione critica C dello spazio dei parametri potrà essere definita come:

$$C = \left\{\mathbf{x} : \frac{L(\mathbf{x};\theta_1)}{L(\mathbf{x};\theta_0)} \geq k_\alpha\right\}. \tag{1.13}$$

Analogamente, la regione di accettazione può essere definita come

$$A = \left\{\mathbf{x} : \frac{L(\mathbf{x};\theta_0)}{L(\mathbf{x};\theta_1)} > k_\alpha\right\}. \tag{1.14}$$

La dimostrazione del lemma, peraltro non difficile, va oltre i nostri scopi, pertanto rinviamo il lettore interessato a un testo di statistica matematica (p.e. Orsi, 1985, p. 410 sgg;), oltre che, ovviamente, all'articolo originale di Neyman e Pearson (1933).

ESEMPIO 1.1

Chiariamo quanto detto con un esempio. Supponiamo che la funzione di densità di probabilità in gioco sia la funzione di Poisson (vedi sopra, la 1.2). Supponiamo anche che precedenti ricerche abbiano dimostrato che gli infortuni sul lavoro in una industria si distribuiscono secondo una poissoniana con $\lambda = 0{,}12$. Ora, secondo i sindacati nell'ultimo anno l'azienda ha risparmiato sulla sicurezza, portando quindi il valore di λ a 0,18. Dobbiamo allora verificare le seguenti ipotesi statistiche:

$$H_0 : \lambda = \lambda_0 = 0,12 \tag{1.15}$$

$$H_1 : \lambda = \lambda_1 = 0,18 \tag{1.16}$$

Immaginiamo che vengano rilevati gli infortuni avvenuti nel corso di questo ultimo anno. Tenuto conto della (1.2) e della (1.5), la nostra funzione di verosimiglianza sarà uguale a

$$L(\mathbf{x};\lambda) = \frac{\lambda^{\sum x_i}}{\prod x_i !} e^{(-n\lambda)}. \tag{1.17}$$

Vediamo allora di calcolare il rapporto di verosimiglianza:

$$\frac{L(\mathbf{x};\lambda_1)}{L(\mathbf{x};\lambda_0)} = \frac{\lambda_1^{\sum x_i}}{\prod x_i !} \frac{\prod x_i !}{\lambda_0^{\sum x_i}} \frac{e^{(-n\lambda_1)}}{e^{(-n\lambda_0)}}. \tag{1.18}$$

Noi però conosciamo i valori di λ_0 e λ_1, rispettivamente 0,12 e 0,18, e possiamo sostituirli nella 1.18. Sulla base della 1.13, possiamo allora impostare la disuguaglianza tale per cui il valore del rapporto sia maggiore o uguale a una costante *c*:

$$\left(\frac{0,16}{0,18}\right)^{\sum x_i} e^{(-0,8n)} \geq c. \tag{1.19}$$

Passando al logaritmo, otteniamo:

$$\sum x_i \log 2 - 0,8n = \sum x_i 0,6931 - 0,8n \geq c. \tag{1.20}$$

Di qui:

$$\sum x_i \geq \frac{\log c + 0,8n}{0,6931} = k_\alpha. \tag{1.21}$$

Si osservi che, come detto sopra, il problema della verosimiglianza è quello di determinare la probabilità che i valori che costituiscono il vettore della variabile casuale campionaria **x** assumano i valori particolari (x_1, x_2, …, x_n), per i quali vale il parametro θ. Possiamo allora sostituire a $\sum x_i$ la $\sum X_i$ e determinare

$$p\left(\sum x_i \geq \frac{\log c + 0,8n}{0,6931} = k_\alpha\right). \tag{1.22}$$

Ora noi sappiamo che questa probabilità deve essere uguale ad α, che abbiamo già fissato a 0,05. Pertanto, una volta stabilita la grandezza del campione (n), risolvendo la (1.21) saremo in grado di conoscere il valore di c. Supponiamo così di avere $n = 20$: c dovrà allora essere uguale a 0,465.

1.4. La verifica della significatività dell'ipotesi nulla (VeSN)

Il lettore si starà chiedendo cosa c'entra tutto questo con la VeSN, per come è stata da noi presentata all'inizio del nostro capitolo. Riprendiamo allora lo schema del processo decisionale, quello in base a cui schema si afferma che il processo inferenziale, in genere detto "fisheriano" richiede i seguenti passaggi: (i) assegnare un valore di probabilità ad α (usualmente 0.05); (ii) assumere che H_0 sia vera; (iii) determinare la probabilità associata ai dati osservati ammesso che H_0 sia vera [cioè, p (dati | H_0)]; (iv) se $p > \alpha$, accettare H_0; (v) altrimenti, rifiutare H_0 e accettare H_1.

Cosa c'è di genuinamente fisheriano in questo schema? Certo non il concetto di due ipotesi contrapposte, nulla e sostantiva, anche se in un certo senso si possono ritenere implicite nel discorso di Fisher. Fisheriano è invece il fatto di affidarsi esclusivamente al valore di probabilità associato ai dati, posto che l'ipotesi sulla non deviazione della forma della loro distribuzione sia corretta, per accettare questa ipotesi o rifiutarla. In realtà, Fisher non dice che dobbiamo cercare di accettare l'ipotesi nulla (se vogliamo usare questa terminologia), ma ci invita direttamente a *verificare* l'ipotesi sostantiva, sulla base della probabilità dell'area di rifiuto.

In questo senso è vero che lo schema decisionale è fisheriano. Il fatto è che la tabella della decisione statistica con questo schema non c'entra nulla, mentre è chiaramente ispirato direttamente da Neyman e Pearson. Qui infatti il compito è diverso: individuare l'area critica per cui si abbia la massima potenza associata all'accettazione dell'ipotesi sostantiva. E come è facile capire, si tratta di due approcci sostanzialmente diversi.

Può essere utile cercare di fornire uno schema che metta in evidenza le principali differenze tra i due approcci. Seguendo in parte Huberty (1993), il lettore può utilmente servirsi della Tabella 1.1.

Nell'approccio AFA, non si ha quindi un giudizio di tutto o di nulla: si ha una considerazione pragmatica della situazione nella sua complessità, che può portare a tre principali esiti: (i) l'accettazione dell'ipotesi sostantiva, (ii) la verifica della necessità di aumentare l'area critica, ad esempio aumentando la grandezza del campione, o (iii) l'accettazione dell'ipotesi nulla (cfr. Neyman, 1950). Tutto ciò dà al ricercatore non solo una maggiore libertà intellettuale di affrontare i dati, ma gli consente anche di padroneggiare situazioni spesso assai complesse nella loro interezza, senza essere imbrigliato da rigidi schematismi, o, per usare l'espressione di Gigerenzer (1998), "rituali di calcolo".

Fisher (PVA)	**Neyman-Pearson (FAA)**
Approccio a probabilità variabile	Approccio ad α prefissato
Test di significatività	Test d'ipotesi
Porre H_0	Porre H_0 e H_1
Specificare la statistica da utilizzare (T) e la sua distribuzione	Specificare la statistica da utilizzare (T) e la sua distribuzione
Raccogliere i dati e calcolare il valore di T	Specificare l'errore di primo tipo α e determinare la regione critica C (regione estrema della distribuzione)
Determinare il valore di p (probabilità condizionale associata ai dati osservati, posto che H_0 sia vera)	Raccogliere i dati e calcolare il valore di T
Rifiutare H_0 se p è piccolo, altrimenti tollerare H_0	Rifiutare H_0 a favore di H_1 se T si trova nella regione d'esclusione della distribuzione, altrimenti tollerare H_0
Concludere l'esperimento	Ripetere se possibile su nuovi campioni

Tabella 1.1 – *Differenze tra l'approccio fisheriano e quello di Neyman-Pearson*

Si osservi che questo ha portato a un notevole utilizzo dell'AFA in applicazioni della statistica anche lontane dalla ricerca scientifica, e in particolare nelle applicazioni industriali, particolarmente nel controllo di qualità.

Un esempio classico è quello della cosiddetta "sberlatura" nell'industria farmaceutica. Il prodotto, ad esempio compresse, viene trasportato su un nastro davanti all'operaio che effettua il controllo, e che lo vede illuminato a luce radente contro uno sfondo adatto. In questo caso, la produzione si sviluppa in continuità, e il controllo deve servire a eliminare le compresse difettose. È evidente che l'interesse dell'industria (e dei consumatori) non è tanto quello di eliminare compresse prive di difetti (falsi allarmi, errori di I tipo), quanto quello di evitare che entrino in commercio com-

presse difettose (omissioni, errori di II tipo). Periodicamente vengono effettuati dei campionamenti, e se il livello α può essere relativamente elevato (anche 0,1), qui è il livello β che va tenuto molto basso, 0,01, o anche e di molto inferiore, se il difetto può portare a danni alla salute del consumatore. Un risultato significativo in uno dei campionamenti porta di conseguenza al blocco della produzione, e all'individuazione della causa che ha portato alla comparsa del difetto.

Come si vede, i due approcci sono sostanzialmente diversi. Come tutto ciò sia stato fuso nell'immaginario collettivo in un unico approccio, è quanto vedremo nel prossimo capitolo.