

LA VALUTAZIONE DEI GIUDICI SENSORIALI: ECCO COSA DICE BSS 2.0

Valutare i giudici a ogni test per eliminare quelli che lo rendono inattendibile e indurli al miglioramento mediante un aiuto concreto e la gratificazione che genera motivazione

Se un test è progettato bene e i dati sono elaborati con tecniche statistiche opportune, il suo valore è correlato in modo diretto all'abilità dei giudici che l'hanno eseguito. Questo è un postulato noto a ogni sensorialista, tanto che in passato stuoli di ricercatori si sono prodigati nel mettere a punto metodi per individuare super assaggiatori e, ancora oggi, c'è chi non ha abbandonato l'obiettivo. Ma è una causa persa in partenza. Non solo perché i super assaggiatori sono per lo più inutili in quanto non rappresentano il percepito comune, ma anche perché, come tutti i loro colleghi normodotati, sono estremamente variabili nell'attribuire l'intensità di una percezione a seconda del loro stato psicofisico.

IL VALORE DELL'ASSAGGIO

Oggi succede che chimici e fisici guardino con aria di superiorità all'analisi sensoriale, considerandola meno attendibile di quanto ottengono dalle loro sofisticate attrezzature, ma un tempo fecero importanti scoperte proprio utilizzando i propri organi di senso.

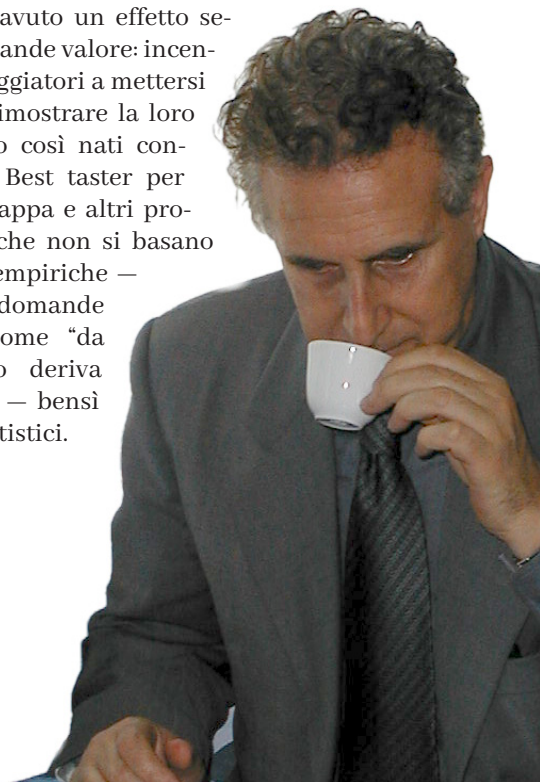
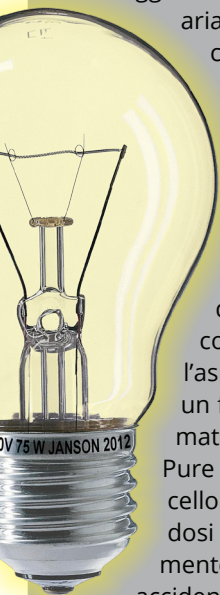
Constantine Fahlberg, un ricercatore che lavorava intorno al catrame, scoprì casualmente la saccarina leccandosi le dita e trovandole incredibilmente dolci, così come successe a un suo collega con l'aspartame e a uno studente, che cercava un farmaco contro la febbre, con il ciclamato di sodio.

Pure la velenosità del pitoui testanera, un uccello che acquisisce tale caratteristica cibandosi di insetti velenosi, fu scoperta casualmente da un ornitologo che si contaminò accidentalmente la mucosa del cavo orale.

Carl Wilhelm Scheele, scopritore di molti elementi tra cui l'acido tartarico, assaggiava abitualmente tutto il materiale di indagine e, ancora oggi, il test più veloce per distinguere il salgemma dalla silvite, apparentemente molto simili, si basa sulla percezione gustativa.

Che un assaggiatore sia affamato o sazio, sereno o irritato, ansioso o euforico, stanco o riposato, tenderà ad attribuire diversi livelli di intensità alla percezione di uno stesso stimolo, per non parlare della variabilità ancora più rilevante che manifesta quando gli viene posta una domanda relativa al valore edonico di un oggetto.

Va da sé che, con l'avvento di nuove tecniche statistiche e dell'informatica che le supporta, i sensorialisti hanno cambiato filosofia, impegnandosi sempre meno nella selezione a priori di giudici validi e sempre più nei metodi che consentono la valutazione del giudice a ogni test che compie, soprattutto nel caso di test ad alta utilità informativa in cui i descrittori sono decine. Questo non solo consente ai panel leader di poter eliminare i giudici in giornata no migliorando così l'attendibilità dei test, ma anche di scoprire dove si collocano le carenze di ogni assaggiatore e quindi di aiutarlo a diventare più bravo prova dopo prova. I progressi compiuti nelle tecniche di valutazione dei singoli componenti del panel hanno avuto un effetto secondario di grande valore: incentivare gli assaggiatori a mettersi in gara per dimostrare la loro bravura. Sono così nati concorsi come i Best taster per vino, caffè, grappa e altri prodotti ancora che non si basano più su prove empiriche — o peggio su domande impossibili come “da quale vitigno deriva questo vino?” — bensì su modelli statistici.



COME SI VALUTANO I GIUDICI

Un giudice compie un buon lavoro quando discrimina i prodotti in analisi dimostrando di saper attribuire i medesimi valori a un campione assaggiato in momenti diversi e di rendere validabile il test in quanto le sue valutazioni sono collimanti con i colleghi.

Vediamo quindi ciò che statisticamente viene considerato.

RIPETIBILITÀ

È la capacità del giudice di valutare lo stesso campione in due momenti diversi assegnando a ogni descrittore valori analoghi. Questa è una caratteristica fondamentale determinata facendo assaggiare uno o più campioni almeno due volte.

COLLIMAZIONE

È la capacità del giudice di dare, per ogni descrittore e per ogni campione, valori simili a quelli attribuiti dai suoi colleghi del panel. La collimazione nei giudizi tra i diversi giudici è alla base del calcolo dell'attendibilità del test, per cui la prova di analisi sensoriale è sempre, o quasi, preceduta da un assaggio di taratura.

DISCRIMINAZIONE

È la capacità del giudice di percepire le differenze tra campioni e di restituirle utilizzando l'intera scala di misura, consentendo quindi una buona discriminazione tra i campioni. Qualsiasi giudice che si riduca automaticamente l'intervallo di scala ha buone probabilità di ottenere una valutazione positiva sia sotto il profilo della collimazione, sia della ripetibilità, ma non è utile al test in quanto non mette in evidenza le differenze esistenti tra campioni.

Attraverso l'attribuzione di pesi specifici per ognuna delle caratteristiche citate si calcola l'indice di efficacia di ogni giudice. Bss 2.0 è impostato per dare un peso di 0,4 alla ripetibilità, 0,25 alla collimazione e 0,35 alla discriminazione (suddivisa in discriminazione tra campioni, discriminazione rispetto ai colleghi, discriminazione rispetto allo storico del Centro Studi Assaggiatori), ma consente di variare le impostazioni a seconda delle necessità.

LA VALUTAZIONE IN PRATICA CON BSS 2.0

Consideriamo un test descrittivo ad alta utilità informativa (Trialtest plus) con 33 descrittori (dei quali 26 oggettivi e 7 edonici) eseguito da 8 giudici su 7 campioni. Eseguendo l'elaborazione della matrice — che nel caso contiene 1.848 dati — il software restituisce la valutazione dei giudici attraverso un grafico o una tabella in cui compaiono i giudici identificati dal codice che li distingueva nel panel, i parametri con i quali sono stati valutati e i valori ottenuti in base alla performance dimostrata.

Dalla *tabella 1* possiamo renderci conto che il gruppo ha lavorato molto bene (in effetti l'attendibilità delle mediane raggiungeva quasi il 95%) e che solo il giudice F è risultato insufficiente, anche se di poco, in ripetibilità. Ma addentriamoci un poco tra i numeri per scoprire qualcosa di più sul comportamento degli altri giudici. Se consideriamo i giudici A e B possiamo notare che il secondo ha un indice di efficacia superiore al primo. Su questo ha sicuramente influito l'ottima ripetibilità. Fantastico se questa performance non fosse dovuta a un vizio di fondo: il giudi-

GIUDICE	RIPETIBILITÀ	COLLIMAZIONE	DISCRIMINAZIONE CAMPIONI	DISCRIMINAZIONE STORICA	DISCRIMINAZIONE PANEL	EFFICACIA
A	7,42	6,72	8,00	5,00	9,00	7,25
B	9,60	8,44	3,60	9,00	9,00	8,29
C	7,16	7,81	6,00	6,00	9,00	7,22
D	7,20	8,10	6,40	4,00	9,00	7,16
E	6,58	5,83	8,40	7,00	10,00	7,05
F	5,87	7,08	4,40	6,00	7,00	6,08
G	8,40	6,57	2,80	7,00	7,00	6,82
H	7,69	7,18	8,40	7,00	10,00	7,83

Tabella 1

LE CARATTERISTICHE DEI GIUDICI

I test di laboratorio (così chiamati per distinguerli da quelli sui consumatori che seguono logiche diverse e non sono svolti da giudici formati) necessitano, soprattutto quando si tratta di test ad alta utilità informativa e quindi molto impegnativi, di giudici distinti da:

- **forte motivazione**, in quanto l'analisi che devono compiere richiede la descrizione del prodotto sotto diversi punti di osservazione, pone molte domande e impone di entrare nei dettagli della percezione. Risulta di conseguenza molto faticosa e la motivazione psicologica si palesa come il migliore antidoto alla stanchezza;
- **notevole onestà intellettuale**, per evitare di compensare con l'astuzia il senso di smarrimento che si può provare nell'incisione sul valore da attribuire e sulla registrazione delle faticose percezioni olfattive, ma anche per evitare di scrivere ciò che si ricorda anziché registrare quello che si sente;
- **forte tendenza all'analisi**, per poter scomporre gli stimoli unitari in percezioni distinte e quindi poter assegnare un valore numerico in funzione dei descrittori proposti, oltre a individuarne di nuovi;
- **sicurezza**, per poter essere rapidi nell'attribuzione dei giudizi e nell'estrazione di nuovi descrittori;
- **notevole capacità descrittiva**, per poter attribuire velocemente etichette semantiche agli stimoli percepiti;
- **buona memoria**, per poter disporre di uno spazio campionario sufficientemente ampio entro il quale collocare con immediatezza il campione per l'aspetto considerato, ma anche per ricordare con facilità le percezioni senza dover assaggiare troppe volte il campione in analisi;
- **ottimo utilizzo delle scale**, senza essere affetti da tendenze minimaliste, massimaliste o pavidie con conseguente distorsione dei giudizi verso valori bassi, alti o mediani.





ce B non ha percepito — o comunque non ha restituito — differenze sui campioni, tanto da meritare un misero 3,60 in “discriminazione campioni”. Un panel leader attento farà tesoro di questa osservazione verificando se rimane abituale, perché potrebbe trattarsi di un giudice dotato di scarsa sensibilità oppure (e più probabilmente) di un valutatore affetto da insicurezza e quindi incline a mantenersi sui valori mediani della scala. Interessante a questo proposito confrontare la prestazione del giudice B con il giudice E che manifesta un’insufficienza nella collimazione, ma un alto valore nella discriminazione tra campioni. È quest’ultimo un assaggiatore che potremmo definire coraggioso, perché non ha assolutamente problemi a esprimere le differenze che percepisce tra i campioni, ma non è molto ripetibile, perché le trova anche sullo stesso campione assaggiato in momenti diversi.

Ma torniamo al giudice F che, non avendo raggiunto la sufficienza nella ripetibilità, va eliminato. Bss 2.0 consente di eseguire l’operazione direttamente dal pannello di controllo, quindi senza dover intervenire sulla matrice, e in cinque secondi ci restituisce la nuova situazione.

Come si può vedere in *tabella 2* l’eliminazione del giudice ha comportato il cambiamento dei valori di tutti i parametri per i quali c’è una relazione tra i diversi giudici, mentre quelli assoluti, quali la ripetibilità, sono rimasti invariati.

I primi tre parametri di valutazione dei giudici sono dunque più importanti, tanto che nella determinazione dell’indice di efficacia pesano l’80%. La discriminazione storica e la discriminazione sul panel pesano appena il 10% cadauno, ma sono di notevole interesse per il panel leader. Mediante il primo può infatti rendersi conto di come sta operando il panel rispetto a quello storico del Centro Studi Assaggiatori e quindi pianificare, nel caso si dimostri necessario, determinate azioni formative. La discriminazione sul panel, in cui ogni giudice viene paragonato ai colleghi attraverso l’uso della scala, risulta utile per rilevare comportamenti anomali di componenti del gruppo per procedere alla loro analisi e alla messa in opera di opportune azioni correttive.

LUIGI ODELLO

GIUDICE	RIPETIBILITÀ	COLLIMAZIONE	DISCRIMINAZIONE CAMPIONI	DISCRIMINAZIONE STORICA	DISCRIMINAZIONE PANEL	EFFICACIA
A	7,42	7,55	8,80	5,00	10,00	7,68
B	9,60	8,62	5,20	9,00	9,00	8,58
C	7,16	8,03	6,40	6,00	10,00	7,43
D	7,20	8,26	7,20	4,00	10,00	7,42
E	6,58	6,63	8,40	7,00	10,00	7,25
G	8,40	6,82	4,00	7,00	7,00	7,06
H	7,69	7,72	8,80	7,00	10,00	8,03

Tabella 2