

Ricerca di soglie critiche per un test su matrici di confusione

Eugenio Brentari[§]

Elena Poli[§]

***Summary:** This paper will discuss misclassification errors induced by human judgement. Facing different classifications coming from different trials, often researchers have to evaluate the goodness of classification proposed. Particularly, we propose an interesting application concerning olfactory measurement. In that field the accuracy of a confusion matrix is crucial, because the classification suggested by a panel of experts is collected by means of a confusion matrix. The analysis of a confusion matrix we will perform in this study serves both to evaluate skills of judges, of whom the panel is composed, and to evaluate the discriminant capability of the element used to make the experiment. This work will focus in establishing critical thresholds for the test thorough which we asses not random classification.*

***Keywords:** Confusion Matrix, Misclassification, Critical Thresholds.*

1. Introduzione

Nell'ambito di indagini economiche e sociologiche nasce spesso l'esigenza di classificare eventi, aleatori o meno, in maniera efficiente tramite tabelle a più vie (Beckwith e Morrison, 1977).

Si consideri un esperimento il cui esito può essere classificato in k categorie, esaustive e mutuamente esclusive, con probabilità date dalla legge di probabilità $p(\cdot)$.

[§] Dipartimento Metodi Quantitativi – Università degli Studi di Brescia – Contrada S. Chiara, 50, 25100 BRESCIA (e-mail: brentari@eco.unibs.it; polielen@eco.unibs.it).

Pur essendo il lavoro frutto di una stretta collaborazione tra gli autori dei quali riflette il pensiero comune, sono ascrivibili a Eugenio Brentari i paragrafi 2 e 3 e a Elena Poli i paragrafi 1 e 4.

Innanzitutto, si cercherà di dare risposta alla domanda su come valutare la classificazione di un evento sulla base dei giudizi umani, in presenza di disaccordo (*misclassification*). Idealmente, dovrebbe esserci una sola classificazione, quella reale, per il fenomeno oggetto di analisi (*recall*), ci sono tuttavia diverse intuizioni, o giudizi individuali, che inducono diverse classificazioni.

Lo strumento statistico di partenza è rappresentato dalla matrice di confusione, che è la matrice contenente le informazioni circa lo stato della realtà e la classificazione ottenuta; in questa matrice le colonne rappresentano la ripartizione per classi indotta dall'esperimento casuale e un dato è assegnato ad una colonna se è classificato come appartenente a tale classe. Le righe rappresentano, invece, le classi reali, si ha quindi che un dato viene rappresentato in una particolare riga se appartiene alla classe corrispondente. Una formalizzazione generale di una matrice di questo tipo è proposta nella Tabella 1.

Tabella 1. Matrice di confusione $k \times k$

		Classificazione				Totale
		1	2	...	k	
Eventi reali	1	n_{11}	n_{12}		n_{1k}	$n_{1\cdot}$
	2	n_{21}	n_{22}		n_{2k}	$n_{2\cdot}$
	\vdots					
	k	n_{k1}	n_{k2}		n_{kk}	$n_{k\cdot}$
Totale		$n_{\cdot 1}$	$n_{\cdot 2}$		$n_{\cdot k}$	n

Una matrice di confusione viene utilizzata per effettuare un *check* sull'accuratezza della classificazione. In generale, quando si considerano esperimenti di classificazione, il modo più immediato per valutare la bontà della classificazione, o più in generale del sistema di classificazione, è effettuare un conteggio del numero di dati classificati in maniera corretta o errata. Questo fornisce una prima, sommaria impressione sulla bontà della classificazione.

Con riferimento all'esemplificazione proposta nella Tabella 1, valori diversi da 0 al di fuori della diagonale principale denunciano errori di classificazione. Col termine inglese *accuracy* si indica la proporzione del numero totale di casi classificati in maniera corretta, data dal rapporto tra la traccia della matrice ed n . Naturalmente, la classificazione è tanto migliore quanto più il valore di *accuracy* si avvicina ad 1.

Una casistica di particolare interesse si ha in presenza di una tabella 2×2 dove due eventi osservabili, si supponga A e \bar{A} , danno luogo ad una classificazione del tipo successo o insuccesso. Dette $p(x)$ e $p(y)$ le due

distribuzioni di probabilità relative, rispettivamente, alle due distribuzioni indotte dagli eventi A e \bar{A} , al fine di condurre considerazioni di tipo probabilistico (per esempio ottenere delle stime o condurre verifiche di ipotesi) assume importanza la valutazione della distribuzione congiunta $p(x, y)$.

Di seguito si proporrà un'interessante applicazione di analisi sensoriale in ambito olfattometrico. In questo campo, esperimenti ripetuti su un *panel* di giudici implicano necessariamente l'intervento del giudizio individuale e fanno emergere, di conseguenza, un problema di corretta classificazione. In particolare, diviene necessario valutare l'affidabilità dei giudici e quindi la bontà della classificazione da essi proposta, dal momento che questa può inficiare i risultati dell'esperimento di analisi sensoriale che si intende condurre. In quest'ottica a partire da una matrice di confusione è possibile calcolare l'*accuracy* della classificazione proposta dai giudici sottoponendo a verifica l'ipotesi che essi forniscano risposte in modo casuale contro l'alternativa che essi siano effettivamente in grado di discriminare in maniera efficiente le diverse situazioni sottoposte al loro giudizio.

L'obiettivo è l'individuazione di soglie critiche che delimitino la regione di rifiuto per test di questo tipo e forniscano allo sperimentatore un semplice criterio operativo e comportamentale per prendere le opportune decisioni.

2. La definizione del problema

L'analisi sensoriale, in quanto disciplina rivolta alla valutazione delle caratteristiche sensoriali di un oggetto tramite gli organi di senso, ponendo al centro dell'attenzione la percezione delle qualità, per esempio di un alimento, rappresenta un ambito naturale in cui emerge un problema di valutazione della classificazione proposta dal giudizio umano. Infatti le aziende interessate a una valutazione delle caratteristiche organolettiche di un proprio prodotto si affidano a un *panel* di giudici perché questi esprimano un parere riguardo al prodotto stesso, anticipando o interpretando il giudizio del consumatore. L'olfattometria, rappresenta quella particolare branca dell'analisi sensoriale che si limita alla considerazione delle percezioni olfattive.

Punto di partenza del presente lavoro è un progetto di ricerca varato dal Centro Nazionale Studi Tartufo che prevede lo sviluppo di un particolare strumento: lo *Sniffer*, ideato dal Centro Studi e Formazione Assaggiatori (Odello, 2004) e realizzato dall'Istituto Nazionale di Fisica della Materia, unità dell'Università degli Studi di Brescia, diretta da G. Sberveglieri del Dipartimento di Chimica e Fisica della Facoltà di Ingegneria che ha curato anche lo sviluppo di un sistema olfattivo artificiale, chiamato naso elettronico (*e-nose*) che individua e identifica un vasto campo di molecole organiche e inorganiche (Pardo *et al.*, 2000; Pardo e Sberveglieri, 2004).

Lo *Sniffer* ha la proprietà di portare contemporaneamente al naso di nove giudici quantità note di molecole in volumi noti di aria. Questo consente di ottimizzare il processo di addestramento dei giudici assaggiatori (un *panel* di esperti capaci di riconoscere il tipo di sostanza che viene fatta loro esaminare) e, parallelamente, un'analisi più accurata del profilo olfattivo del prodotto oggetto di valutazione. Non è casuale che lo *Sniffer* sia stato ideato proprio per il tartufo, un tubero così prezioso di cui ai giudici non è dato godere fino in fondo, ma solo di compiere una valutazione non distruttiva in cui la parte olfattiva è largamente prevalente. La prima versione dello *Sniffer*, impiegata nel 2002, è stata perfezionata e collaudata nel giugno 2004 ed è pronta per essere impiegata nel piano di ricerca del Centro Nazionale Studi Tartufo.

Con l'uso dello *Sniffer*, che consente di convogliare nelle narici dagli "annusatori", in una fissata unità di tempo (generalmente 40 secondi), volumi di aria noti a concentrazioni note di sostanze volatili, si intende quindi valutare da un canto le soglie di percezione dei singoli giudici e dall'altro le soglie di percezione di sostanze olfattive che possono essere di interesse per la caratterizzazione dei tartufi sulla base dell'origine.

Nell'applicazione, si vengono così a incontrare aspetti prettamente applicativi con problemi di natura statistica.

Si consideri un esperimento nel quale il *panel* di riferimento risulta costituito da un insieme di giudici, ognuno dei quali viene sottoposto a un test nel quale, con l'ausilio dello *Sniffer*, annusa una sostanza gassosa; questa può essere aria pura, oppure una miscela di aria e di un campione (che in seguito verrà indicato come *analita*). Compito del giudice è riconoscere la presenza o l'assenza dell'analita. Chi presiede l'esperimento e sa quando viene immesso l'analita (in accordo con un piano degli esperimenti) deve semplicemente annotare le risposte corrette ed errate fornite dal *tester*.

Per un numero n di prove ripetute sottoposte a un singolo giudice, nelle quali n_p volte l'analita è effettivamente presente mentre n_a volte è assente (naturalmente, $n_p + n_a = n$) i risultati ottenuti possono essere riassunti mediante una matrice di confusione 2×2 del tipo presentato in Tabella 2.

Tabella 2. Matrice di confusione 2×2 in cui si possono classificare gli esiti dell'esperimento in esame

		<i>Parere del giudice</i>		Totale
		<i>rilevato</i>	<i>non rilevato</i>	
<i>Analita</i>	<i>presente</i>	n_{pr}	n_{pn}	n_p
	<i>assente</i>	n_{ar}	n_{an}	n_a
Totale		n_r	n_n	N

I casi favorevoli sono quelli in cui il parere del giudice coincide con la realtà nota a chi verifica l'esito della prova; corrispondono quindi alla somma di n_{pr} (il numero di volte in cui l'analita è presente ed è stato individuato) e di n_{an} (il numero di volte in cui non c'è e correttamente ne è stata riconosciuta l'assenza).

L'*accuracy* in questo caso è data da $\frac{n_{pr} + n_{an}}{n}$.

È necessario interpretare i dati che emergono dalle prove e, in particolare modo, valutare l'affidabilità della classificazione proposta dal singolo giudice e dall'intero *panel*, per poter trarre conclusioni circa due aspetti dell'esperimento.

In primo luogo, trarre delle conclusioni relative al livello di concentrazione dell'analita; è possibile infatti dedurre se il livello di concentrazione è tale da consentirne il riconoscimento da parte del *panel* considerato o alternativamente si può concludere che l'intensità odorosa dello stimolo è troppo bassa per le capacità olfattive dei soggetti sottoposti al test¹. In quest'ultima ipotesi può risultare utile "ritarare" l'esperimento.

Inoltre, per quanto riguarda la qualità dei giudizi del *panel*, è possibile "misurare" quanto sia affidabile il parere espresso dai giudici.

Per entrambi gli aspetti (individuazione della soglia di percezione e selezione dei giudici) si cercherà di individuare un criterio ottimale di decisione.

3. La distribuzione della statistica test

Al termine di esperimenti come quelli descritti nel paragrafo precedente, il ricercatore sulla base dei risultati dell'esperimento (sintetizzati come in Tabella 2) deve esprimere una valutazione degli stessi. In particolare dovrebbe essere in grado di valutare se le risposte fornite nell'esperimento siano attendibili, oppure siano state date in maniera casuale.

Per effettuare questo test risulta utile indagare il modello probabilistico in grado di descrivere al meglio la distribuzione del numero di risposte corrette date dai giudici.

Ogni singola prova dell'esperimento può essere vista come una variabile casuale di Bernoulli, dal momento che genera una popolazione dicotomica

¹ La soglia (o limite) di percezione di un composto odoroso viene definita come quella concentrazione dell'odore di una sostanza nell'aria avvertita da almeno il 50% del gruppo di valutazione (si veda, per esempio, Versini, 2003).

Una definizione di soglia di percezione, attribuita a Bengtsson (1953), fa riferimento a un "fantomatico" intervallo di confidenza. Nel lavoro citato però non vi è traccia di una simile definizione (si veda, Meilgaard *et al.*, 1991).

del tipo successo/insuccesso (classificazione corretta/classificazione errata). La ripetizione dell'esperimento a condizioni invariate, genera una serie di risposte tra loro indipendenti per cui la distribuzione di probabilità di riferimento è una binomiale.

Si possono così individuare due variabili casuali aventi distribuzione binomiale:

- $X_p \sim b(n_p, p_p)$ identifica il numero di successi (classificazioni corrette) in n_p prove, quando la probabilità di successo è p_p ;
- $X_a \sim b(n_a, p_a)$ identifica invece il numero di successi in n_a prove, quando la probabilità di successo è p_a .

L'interesse del ricercatore riguarda il numero complessivo di classificazioni corrette derivante dalle due distribuzioni di probabilità che, per come è costruito l'esperimento, risultano indipendenti.

Qualora la risposta del giudice sia fornita casualmente ($p = p_p = p_a = 0,5$), la distribuzione di probabilità congiunta di X_p e X_a , per la proprietà riproduttiva della distribuzione binomiale (Landenna *et al.*, 1997; Brentari *et al.*, 1999), è a sua volta una binomiale di parametri $n = n_p + n_a$ e p .

Si ha così che la distribuzione di probabilità del numero S di classificazioni corrette effettuate in n prove da un singolo soggetto è una binomiale di parametri n e p ; quindi

$$S \sim b(n, p).$$

Indicato con q il numero di individui che partecipano all'esperimento e ipotizzato che questi non possano influenzarsi durante le prove, il numero complessivo di classificazioni corrette $W = S_1 + S_2 + \dots + S_q$ fornite da tutti i soggetti, sempre per la proprietà riproduttiva, si distribuisce ancora come una variabile casuale binomiale di parametri nq e p . Allora, il numero W di classificazioni corrette fornite da tutti i soggetti sarà così distribuito

$$W \sim b(nq, p).$$

4. Individuazione delle soglie critiche per il test

Il problema che sorge nel gestire un esperimento nell'ambito descritto potrebbe essere, come si è detto, quello di valutare la capacità di discernimento dei giudici (ma un'analoga impostazione potrebbe riguardare la soglia di percezione).

Nel caso in cui non ci sia una particolare abilità olfattiva il parere del *tester* è paragonabile a quello di chi risponde a caso; la probabilità di una classificazione corretta risulta quindi uguale a quella di una classificazione erronea, e quindi pari a $p = 0,5$.

In questo caso può allora essere formulata un'ipotesi da sottoporre a verifica che può essere così sintetizzata $H_0: p = 0,5$. Sotto l'ipotesi nulla così formulata la v.c. numero di classificazioni esatte avrebbe distribuzione binomiale di parametri nq e $p = 0,5$; $W \sim b(nq; 0,5)$.

L'ipotesi $p = 0,5$ può essere sottoposta a verifica contro un'ipotesi alternativa unilaterale che il *panel* sia in grado di discriminare correttamente le diverse situazioni, quindi $H_1: p > 0,5$. Il sistema d'ipotesi è quindi

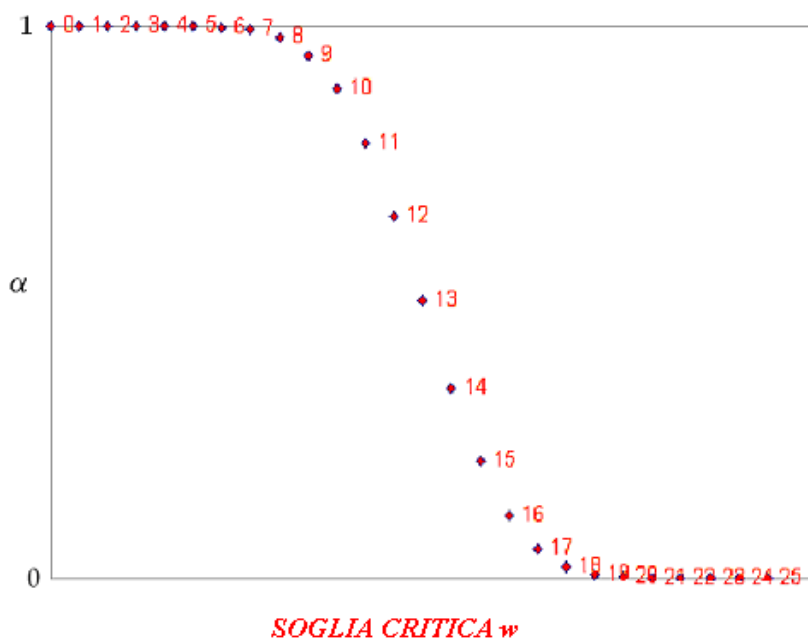
$$H_0: p = 0,5,$$

$$H_1: p > 0,5.$$

Si consideri nuovamente l'esperimento introduttivo che preveda, per esempio, di sottoporre cinque giudici a cinque prove nelle quali tre volte è presente l'analita e assente nelle restanti due.

Si procede determinando per diversi valori del livello di significatività α il valore soglia che individua la regione di rifiuto del test. Nella Figura 1 viene messa in evidenza la relazione tra le soglie della regione di rifiuto dell'ipotesi nulla e il livello di significatività del test.

Figura 1. Livello di significatività α al variare della soglia critica w (caso con cinque giudici per cinque prove)



Per esempio, qualora si fissasse il livello di significatività al 2% la zona di rifiuto risulterebbe essere $R = \{W \geq 19\}$.

Si noti che trovandosi di fronte una variabile casuale discreta può non esistere una zona di rifiuto cui corrisponda il livello di significatività desiderato, infatti nell'esempio precedente il reale livello di significatività è 0,073% e non il 2% prefissato, del resto, tale scelta è in linea con un atteggiamento "conservativo" nei confronti dell'ipotesi nulla che suggerisce di scegliere una zona di rifiuto che riduca la probabilità dell'errore di prima specie piuttosto di una che la aumenti.

Pertanto, la soglia di rifiuto w viene determinata scegliendo il valore maggiore della v.c. W che soddisfa la disuguaglianza

$$P(W \geq w | H_0) \leq \alpha.$$

Una delle conseguenze di tale approssimazione è che per un dato livello di significatività si può individuare la medesima soglia anche in presenza di numerosità nq diverse.

In Appendice viene presentata la tabella relativa alle soglie critiche (Tabella A) per esperimenti con diversa numerosità complessiva ($nq = 10, 11, \dots, 150$), stabilite ai livelli di significatività più comunemente utilizzati ($\alpha = 0,001; 0,005; \dots; 0,1$).

Ovviamente le soglie critiche così costruite restituiscono un test più conservativo rispetto al valore nominale di α . Per questo si è prestata particolare attenzione alla valutazione dello scostamento tra valore effettivo e valore nominale del livello di significatività α . Considerando numerosità campionarie che vanno da 10 a 150, la differenza tra il valore effettivo di α e quello nominale ha presentato valori medi assai ridotti, come evidenziato nella Tabella 3.

Tabella 3 Confronto tra livello di significatività effettivo e nominale

	α nominale					
	0,001	0,005	0,01	0,025	0,05	0,1
Media degli scarti	0,0003	0,0016	0,0029	0,0065	0,0115	0,0207

Appare, infine, ragionevole considerare un test di verifica di ipotesi in cui l'ipotesi alternativa viene specificata diversamente, per esempio, come $p = p_{pr} = p_{an} = 0,8$. Nell'ambito dell'analisi sensoriale si potrebbe pensare di richiedere che un giudice esprima giudizi corretti almeno nell'80% dei casi.

L'espressione di un valore numerico che esprima il concetto di “*panel* valido” (0,8, appunto) consente di prevedere la probabilità con cui si giudicherà un “*panel* valido” oppure “non valido”.

Supponendo di avere un esperimento di 30 elementi (6 giudici per 5 prove) e di desiderare un livello di significatività dell'1%, la soglia critica risulterebbe pari a 22; in questo caso la probabilità dell'errore di seconda specie sarebbe del 12,87% dal momento che

$$\beta = P(W < 22 | nq = 30; p = 0,8) = 0,1287.$$

È noto che, dato un certo campione, la riduzione di un tipo di errore produce l'aumento dell'altro; infatti l'unica via che porta alla riduzione congiunta di α e β è l'aumento della numerosità campionaria.

Tuttavia questa soluzione non è sempre percorribile per ragioni di costo o per altri motivi pratici: nel caso in esame, oltre a un vincolo temporale, esiste il vincolo imposto dal limitato numero di prove cui il naso umano può sottoporsi senza perdere sensibilità.

È necessario quindi scendere a compromessi e considerare l'eventualità di accettare un aumento dell'errore di prima specie qualora quello di seconda specie venga ritenuto eccessivamente elevato.

Se, per esempio, nel caso precedente si considerasse troppo elevata la probabilità dell'errore $\beta = 12,87\%$ si potrebbe aumentare il livello di significatività del test, nell'impossibilità di aumentarne la numerosità. Accettando un livello di significatività α corrispondente al 2,5%, la zona di rifiuto diventerebbe $R = \{W \geq 21\}$ e la probabilità β dell'errore di seconda specie con questa nuova soglia scenderebbe al 6,11%.

5. Conclusioni

Nel presente lavoro si è preso in considerazione il problema rappresentato dagli errori di classificazione indotti dal giudizio umano; infatti, ogni qualvolta sorge l'esigenza di riconoscere e classificare diverse situazioni relative a un insieme di esperimenti ripetuti, nasce al contempo l'esigenza da parte dello sperimentatore di valutare l'affidabilità della classificazione cui si perviene. In particolare, in questo studio si è proposta un'applicazione nell'ambito dell'olfattometria: lo sviluppo del progetto dello *Sniffer*, uno strumento che prevede l'addestramento di un gruppo di giudici affinché siano in grado di distinguere le sostanze odorose.

L'interpretazione dell'accuratezza della classificazione individuata dal *panel* di giudici e proposta in una tabella di confusione diviene cruciale in questo campo con la duplice finalità di valutare l'affidabilità del *panel* e di individuare la capacità discriminante dell'analisi.

In particolare, si pone l'accento sull'individuazione delle soglie critiche del test condotto in relazione all'ipotesi di casualità nella classificazione.

Dal punto di vista pratico, tale approfondimento rende disponibile allo sperimentatore una semplice e ragionevole regola operativa per valutare la bontà dei giudizi.

Al di là dell'ambito di applicazione qui individuato, si rimarca la vastità degli ambiti di interesse per analisi di questo genere. È noto, per esempio, l'interesse generato da situazioni di questo tipo, in ambito medico e biometrico, dove a una coppia di eventi del tipo "farmaco" o "placebo" corrisponde una classificazione degli effetti del tipo esito positivo o meno, secondo una logica che si desidera essere di tipo non casuale.

Ringraziamenti

Gli autori desiderano ringraziare il Centro Studi e Formazione Assaggiatori, l'Istituto Nazionale di Fisica della Materia, unità di Brescia, e il Centro Nazionale Studi Tartufo per aver permesso di seguire lo sviluppo di *Sniffer* che ha originato il presente lavoro.

Appendice

Tabella A *Soglie critiche*

<i>n</i>	α						<i>n</i>	α					
	0,001	0,005	0,01	0,025	0,05	0,1		0,001	0,005	0,01	0,025	0,05	0,1
10	10	10	10	9	9	8	44	33	31	31	29	28	27
11	11	11	10	10	9	9	45	34	32	31	30	29	28
12	12	11	11	10	10	9	46	34	33	32	31	30	28
13	13	12	12	11	10	10	47	35	33	32	31	30	29
14	13	13	12	12	11	10	48	36	34	33	32	31	29
15	14	13	13	12	12	11	49	36	34	34	32	31	30
16	15	14	14	13	12	12	50	37	35	34	33	32	31
17	16	15	14	13	13	12	51	37	36	35	33	32	31
18	16	15	15	14	13	13	52	38	36	35	34	33	32
19	17	16	15	15	14	13	53	39	37	36	35	33	32
20	18	17	16	15	15	14	54	39	37	36	35	34	33
21	18	17	17	16	15	14	55	40	38	37	36	35	33
22	19	18	17	17	16	15	56	40	39	38	36	35	34
23	20	19	18	17	16	16	57	41	39	38	37	36	34
24	20	19	19	18	17	16	58	42	40	39	37	36	35
25	21	20	19	18	18	17	59	42	40	39	38	37	35
26	22	20	20	19	18	17	60	43	41	40	39	37	36
27	22	21	20	20	19	18	61	43	41	41	39	38	37
28	23	22	21	20	19	18	62	44	42	41	40	38	37
29	24	22	22	21	20	19	63	45	43	42	40	39	38
30	24	23	22	21	20	20	64	45	43	42	41	40	38
31	25	24	23	22	21	20	65	46	44	43	41	40	39
32	26	24	24	23	22	21	66	46	44	43	42	41	39
33	26	25	24	23	22	21	67	47	45	44	42	41	40
34	27	25	25	24	23	22	68	48	46	45	43	42	40
35	27	26	25	24	23	22	69	48	46	45	44	42	41
36	28	27	26	25	24	23	70	49	47	46	44	43	41
37	29	27	27	25	24	23	71	49	47	46	45	43	42
38	29	28	27	26	25	24	72	50	48	47	45	44	42
39	30	28	28	27	26	24	73	51	48	47	46	45	43
40	31	29	28	27	26	25	74	51	49	48	46	45	44
41	31	30	29	28	27	26	75	52	50	49	47	46	44
42	32	30	29	28	27	26	76	52	50	49	48	46	45
43	32	31	30	29	28	27	77	53	51	50	48	47	45

n	α						n	α					
	0,001	0,005	0,01	0,025	0,05	0,1		0,001	0,005	0,01	0,025	0,05	0,1
78	54	51	50	49	47	46	115	75	72	71	69	67	65
79	54	52	51	49	48	46	116	76	73	71	70	68	66
80	55	52	51	50	48	47	117	76	73	72	70	68	66
81	55	53	52	50	49	47	118	77	74	73	71	69	67
82	56	54	52	51	49	48	119	77	74	73	71	69	67
83	56	54	53	51	50	48	120	78	75	74	72	70	68
84	57	55	54	52	51	49	121	78	76	74	72	71	69
85	58	55	54	53	51	49	122	79	76	75	73	71	69
86	58	56	55	53	52	50	123	80	77	75	73	72	70
87	59	56	55	54	52	50	124	80	77	76	74	72	70
88	59	57	56	54	53	51	125	81	78	76	74	73	71
89	60	58	56	55	53	52	126	81	78	77	75	73	71
90	61	58	57	55	54	52	127	82	79	78	76	74	72
91	61	59	58	56	54	53	128	82	80	78	76	74	72
92	62	59	58	56	55	53	129	83	80	79	77	75	73
93	62	60	59	57	55	54	130	84	81	79	77	75	73
94	63	60	59	57	56	54	131	84	81	80	78	76	74
95	63	61	60	58	57	55	132	85	82	80	78	76	74
96	64	62	60	59	57	55	133	85	82	81	79	77	75
97	65	62	61	59	58	56	134	86	83	81	79	78	75
98	65	63	61	60	58	56	135	86	83	82	80	78	76
99	66	63	62	60	59	57	136	87	84	83	80	79	76
100	66	64	63	61	59	57	137	87	85	83	81	79	77
101	67	64	63	61	60	58	138	88	85	84	81	80	78
102	67	65	64	62	60	58	139	89	86	84	82	80	78
103	68	66	64	62	61	59	140	89	86	85	83	81	79
104	69	66	65	63	61	60	141	90	87	85	83	81	79
105	69	67	65	64	62	60	142	90	87	86	84	82	80
106	70	67	66	64	62	61	143	91	88	86	84	82	80
107	70	68	66	65	63	61	144	91	88	87	85	83	81
108	71	68	67	65	64	62	145	92	89	87	85	83	81
109	72	69	68	66	64	62	146	93	90	88	86	84	82
110	72	69	68	66	65	63	147	93	90	89	86	84	82
111	73	70	69	67	65	63	148	94	91	89	87	85	83
112	73	71	69	67	66	64	149	94	91	90	87	86	83
113	74	71	70	68	66	64	150	95	92	90	88	86	84
114	74	72	70	68	67	65							

Riferimenti Bibliografici

- Beckwith N.E., Morrison D.G. (1977). Stochastic Interpretation of 2×2 Classification Table. *Journal of the American Statistical Association*, **72**, 358, 303-308.
- Bengtsson K. (1953). *Taste testing as an analytical method statistical treatment of the data*. Wallerstein Communications, Stoccolma.
- Brentari E., Carpita M., Poli E. (1999). *Le variabili casuali*. CLUB, Brescia, (con CD-Rom).
- Do K., McLachlan G. J. (1984). Estimation of Mixing Proportions: A Case Study. *Applied Statistics*, **33**, 2, 134-140.
- Landenna G., Marasini D., Ferrari P.(1997). *Probabilità e variabili casuali*. Il Mulino, Bologna.
- Mailgaard D., Civille G., Carr B.(1991). *Sensory evaluation Techniques*. CRC Press, Boca Raton (Florida).
- Odello L. (2004). Dal mondo dell'assaggio e ... dintorni. *L'assaggio*, **VII**, 7 n.s., 60-61.
- Pardo M., Niederjaufner G, Benussi G., Comini E., Faglia G., Sberveglieri G., Holmberg M., Lundstrom I. (2000). Data Processing enhances the classificatrion of different brands of Espresso coffee with an Electronic Nose. *Sensor and Actuators, B*, **69**, 397-403.
- Pardo M., Sberveglieri G. (2004). Electronic Olfactory Systems Based on Metal Oxide Semiconductor Sensor Arrays. *MRS Bulletin*, to be published provisionally in the October 2004 Special Issue *Novel Materials and Applications of Electronic Noses and Tongues*.
- Teather D. (1974). Statistical Techniques for Diagnosis, *Journal of the Royal Statistical Society*. **137**, 2, 231-244.
- Versini G. (2003). Cosa c'è dietro l'aroma dei vini. *L'assaggio*, **VII**, 2 n.s., 21-27.
- Zenga M. (1968). La distribuzione binomiale bivariata. *Statistica*, 28, 1, 83-101.