5. PROBLEMES D'ESTADÍSTICA I PROBABILITATS

5.1. Introducció

Són molt variats els problemes d'aquesta naturalesa que es poden presentar en la gestió de l'empresa agrària. De fet, la teoria clàssica de probabilitats pot tenir utilitats estimables al món agroalimentari i forestal, com es pot veure als exemples que desenvolupem a continuació per aplicació del conegut Teorema de Bayes[16] i d'altres conceptes de la probabilitat condicionada, els fonaments teòrics dels quals es poden trobar a l'annex núm. 5 del nostre llibre (veure el CD adjunt).

Vegem-los seguidament:

5.2. Primer problema

Doncs bé, el problema que ara se'ns planteja per resoldre, dins l'àmbit del control de qualitat d'un procés agroindustrial, és el següent: "En un celler cooperatiu de la Terra Alta hi ha tres màquines etiquetadores que produeixen, respectivament, el 50%, 30% i 20% de les etiquetes. Els percentatges respectius d'etiquetes defectuoses per a cada màquina són el 3%, 4% i 5%. En certa ocasió, el distribuïdor reclama una partida d'ampolles etiquetades defectuosament. Quina és la probabilitat de què hagi estat feta per cadascuna de les màquines?"

Solució:

Es tenen les següents probabilitats:

$$\begin{cases} P(M_1) = 0'50 \rightarrow P(D|M_1) = 0'03 \\ P(M_2) = 0'30 \rightarrow P(D|M_2) = 0'04 \\ P(M_3) = 0'20 \rightarrow P(D|M_3) = 0'05 \end{cases}$$

Així doncs, la probabilitat de què l'ampolla etiquetada defectuosament provingui de la primera màquina serà:

---

[16] El **teorema de Bayes**, força conegut, és un dels teoremes més emprats a la teoria clàssica de la probabilitat. Descobert per Thomas Bayes és una manera particular de relacionar dues probabilitats per tal de demostrar la relació entre la probabilitat d'un esdeveniment condicionada al succés d'un segon esdeveniment i la probabilitat d'aquest segon esdeveniment condicionada al succés del primer, és a dir, entre P(A/B) i P(B/A).

$$P(M_1|D) = \frac{P(M_1) \times P(D|M_1)}{P(M_1) \times P(D|M_1) + P(M_2) \times P(D|M_2) + P(M_3) \times P(D|M_3)} =$$

$$= \frac{0'50 \times 0'03}{0'50 \times 0'03 + 0'30 \times 0'04 + 0'20 \times 0'05} = \frac{0'015}{0'037} = \frac{15}{37} \cong 41\%$$

De la mateixa manera, la probabilitat de què procedeixi de la segona màquina serà:

$$P(M_2|D) = \frac{0'30 \times 0'04}{0'50 \times 0'03 + 0'30 \times 0'04 + 0'20 \times 0'05} = \frac{0'012}{0'037} = \frac{12}{37} \cong 32\%$$

I per últim, de la tercera màquina es tindrà:

$$P(M_3|D) = 1 - P(M_1|D) - P(M_2|D) = 1 - \frac{15}{37} - \frac{12}{37} = \frac{10}{37} \cong 27\%$$

5.3. Segon problema

"El gerent d'una agrobotiga cooperativa ha observat el capteniment dels seus clients durant un llarg període de temps. Com a conseqüència d'aquest període d'observació afirma que la probabilitat de què un client que entri a la botiga compri una ampolla de vi és del 40%, però dels que compren una ampolla de vi el 50% compren també una ampolla d'oli, i només un 10% compren l'ampolla d'oli quan no han comprat l'ampolla de vi. Es vol obtenir les probabilitats de què els clients compren el següent:

1. Una ampolla de vi i una d'oli.
2. Una ampolla d'oli.
3. Una ampolla de vi o una d'oli.
4. Una ampolla d'oli però no una de vi."

FIG. 9.20. Cooperativa Soldebre SCCL.

Solució:

Considerarem els dos successos bàsics:

V: Compra d'una ampolla de vi.
O: Compra d'una ampolla d'oli.

Sabem, per les dades del problema, que s'acompleix que:

$P(V) = 0'4$
$P(O|V) = 0'5$
$P(O|\overline{V}) = 0'1$

L'espai mostral, per a aquest experiment aleatori, serà el següent:

$$E = \{V \cap O,\ V \cap \overline{O},\ \overline{V} \cap O,\ \overline{V} \cap \overline{O}\}$$

Per a la consecució d'una millor comprensió del problema plantejat, resulta útil fer la representació mitjançant el següent diagrama de Venn-Euler, així:

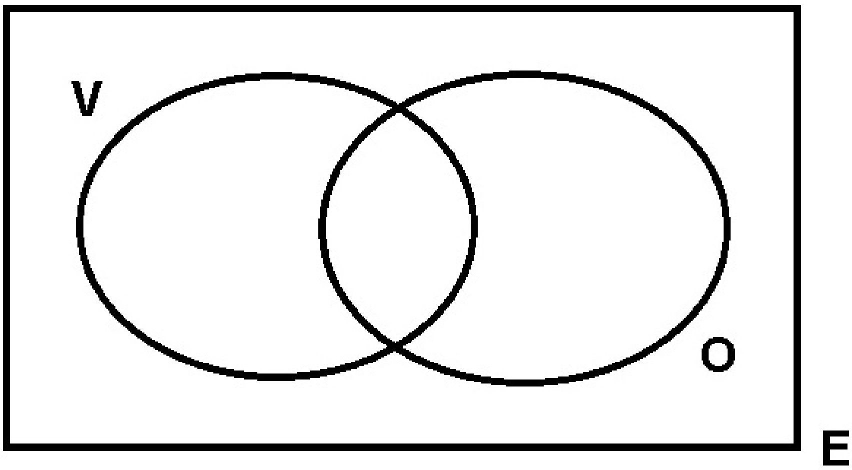


FIG. 9.21. Diagrama de Venn-Euler (I).

Les probabilitats dels successos que ens demanen són:

1. Probabilitat de comprar una ampolla de vi i una d'oli.

$$P(V \cap O) = P(V) \cdot P\,(O|V) = 0'4 \cdot 0'5 = 0'2 = 20\%$$

2. Probabilitat de comprar una ampolla d'oli.

$$P(O) = P[(V \cap O) \cup (\overline{V} \cap O)] = P(V \cap O) + P(\overline{V} \cap O) = 0'2 + 0'06 = 0'26 =$$
$$=26\%$$

ja que,

$$P(\overline{V} \cap O) = P(\overline{V}) \cdot P(O|\overline{V}) = 0'6 \cdot 0'1 = 0'06$$

3. Probabilitat de comprar una ampolla de vi o una d'oli.

$$\begin{aligned} P(V \cup O) &= P[(V \cap O) \cup (V \cap \overline{O}) \cup (\overline{V} \cap O)] = \\ &= P(V \cap O) + P(V \cap \overline{O}) + P(\overline{V} \cap O) = \\ &= 0'2 + 0'2 + 0'06 = 0'46 = 46\% \end{aligned}$$

ja que,

$$P(V \cap \overline{O}) = P(V) \cdot P(\overline{O}|V) = 0'4 \cdot 0'5 = 0'2$$

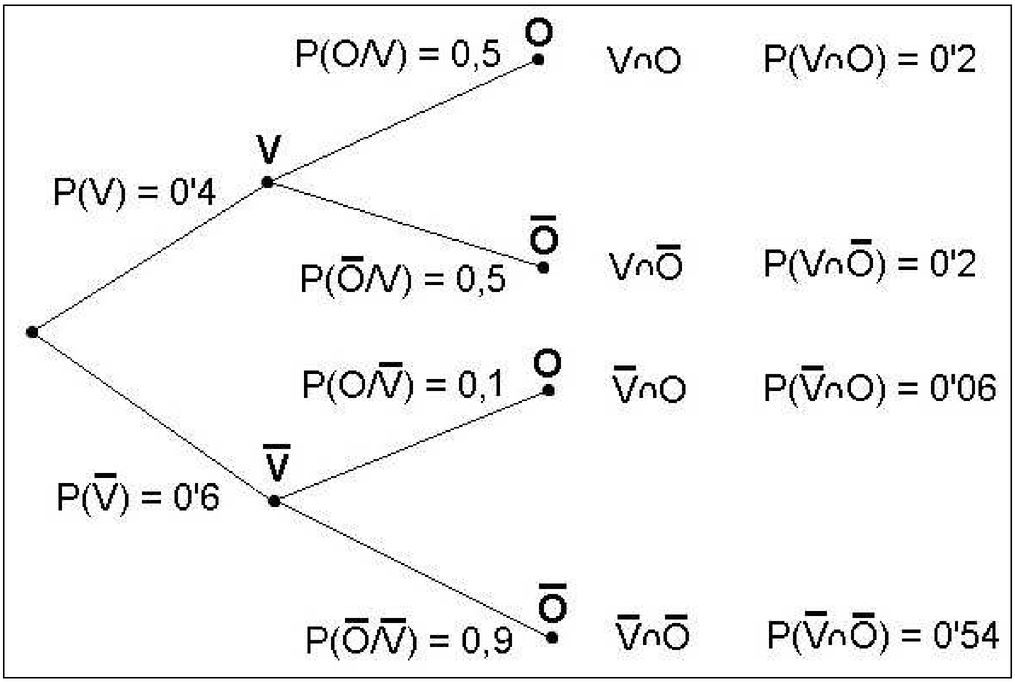
o bé, directament pel teorema de la probabilitat total,

$$P(V \cup O) = P(V) + P(O) - P(V \cap O) = 0'4 + 0'26 - 0'2 = 0'46 = 46\%$$

4. Probabilitat de comprar una ampolla d'oli però no una de vi.

$$P(\overline{V} \cap O) = P(\overline{V}) \cdot P(O|\overline{V}) = 0'6 \cdot 0'1 = 0'06 = 6\%$$

El corresponent arbre de probabilitat o graf arborescent seria, en aquest cas:



I a l'espai mostral total es compleix que:

P(E) = 0'20 + 0'20 + 0'06 + 0'54 = 1'00, com es volia demostrar.

5.4. Tercer problema

"Una almàssera disposa de tres línies de producció que produeixen 1.000, 2.000 i 4.000 ampolles de tres tipus diferents d'oli, respectivament, segons el seu grau d'acidesa. La proporció d'ampolles que no superen el control de qualitat establert a l'efecte és del 1%, 2% i 3%, respectivament. Calcular:

a) La probabilitat de què una ampolla de l'empresa no superi el control de qualitat exigible.

b) Si s'observa una ampolla qualsevol i supera el control de qualitat, quina és la probabilitat de què hagi estat fabricada en la tercera línia de producció?"

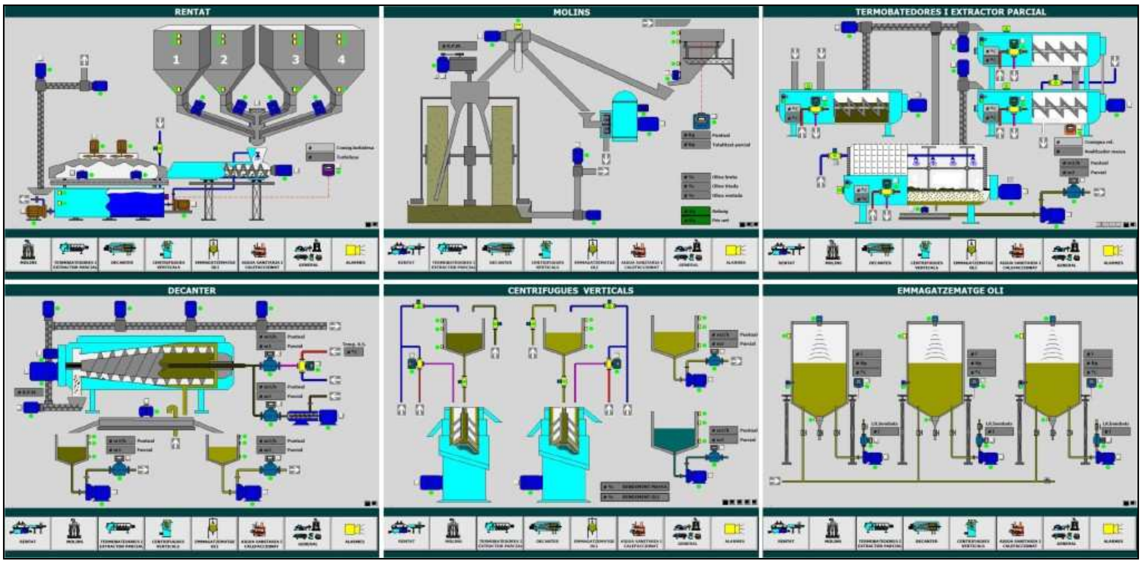


FIG. 9.22. Quadre sinòptic d'una almàssera.

Solució:

a) Evidentment, qualsevol ampolla ha estat fabricada en la 1a, 2a o 3a línia i solament en una d'aquestes. Anomenem F_1, F_2, i F_3 al succés "l'ampolla ha estat fabricada en la 1a, 2a o 3a línia respectivament".

Anomenem C al succés "supera el control de qualitat del producte de l'empresa" i $\overline{C}$ serà el seu succés contrari o complementari.

Pel teorema de la probabilitat total es tindrà que:

$$P(\overline{C}) = P(F_1) \cdot P(\overline{C} \mid F_1) + P(F_2) \cdot P(\overline{C} \mid F_2) + P(F_3) \cdot P(\overline{C} \mid F_3) =$$

$$= \frac{1.000}{7.000} \cdot 0'01 + \frac{2.000}{7.000} \cdot 0'02 + \frac{4.000}{7.000} \cdot 0'03 = \frac{1}{700} + \frac{4}{700} + \frac{12}{700} = \frac{17}{700} \approx 2'43\%$$

b) Ara es tractarà de calcular la següent probabilitat:

$$P(F_3 | C) = \frac{P(F_3) \cdot P(C | F_3)}{P(C)} = \frac{P(F_3) \left[1 - P(\overline{C} | F_3)\right]}{1 - P(\overline{C})} =$$

$$= \frac{\frac{4}{7}(1 - 0'03)}{1 - \frac{17}{700}} = \frac{\frac{4}{7} \cdot 0'97}{\frac{683}{700}} = \frac{388}{683} \approx 56'81\%$$

Això s'ha resolt aplicant la definició de probabilitat condicionada, i emprant la propietat general $P(\overline{S}) = 1 - P(S)$, i l'apartat a) d'aquest mateix exercici.

5.5. Quart problema

"En una comarca ebrenca, la probabilitat de què una explotació porcina contamini, si hi ha una llei de protecció ecològica, és de l'1%. La probabilitat de què es promulgui pel Parlament de Catalunya una llei de protecció ecològica és del 50%, i la probabilitat de què una explotació ramadera d'aquestes característiques contamini el medi ambient és del 10%, segons els estudis realitzats prèviament. Calcular:

a) La probabilitat de què l'explotació no contamini i hi hagi llei de protecció ecològica.
b) La probabilitat de què contaminant l'explotació, hi hagi llei de protecció ecològica.
c) La probabilitat de què no havent-hi llei de protecció ecològica, l'explotació no contamini.
d) La probabilitat de què havent-hi llei de protecció ecològica, l'explotació no contamini."

FIG. 9.23. Granja de porcs a les comarques de l'Ebre.

Solució:

Anomenem ara L al succés "es promulga llei de protecció ecològica", i C al succés "l'explotació porcina contamina". Les dades del problema són les següents:

$$P(C|L) = 0'01 \quad ; \quad P(L) = 0'50 \quad ; \quad P(C) = 0'10$$

També resulta útil aquí fer la representació en diagrama de Venn-Euler, així:

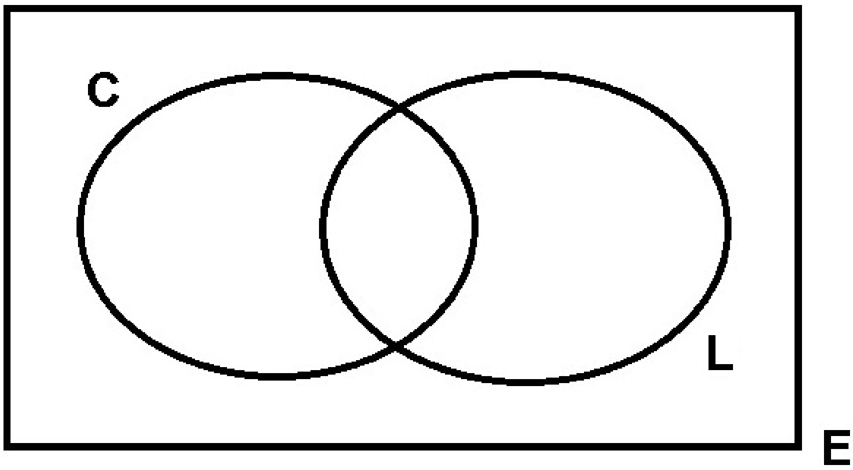


FIG. 9.24. Diagrama de Venn-Euler (II).

Ara resoldrem separadament les diferents qüestions plantejades:

a)

$$P(\overline{C} \cap L) = P(L) \cdot P(\overline{C}|L) = 0'5 \cdot [1 - P(C|L)] = 0'5 \cdot [1 - 0'01] =$$

$$= 0'5 \cdot 0'99 = 0'495 = 49'5\%$$

b)

$$P(L|C) = \frac{P(L \cap C)}{P(C)} = \frac{P(L) - P(\overline{C} \cap L)}{P(C)} =$$

$$= \frac{0'5 - 0'495}{0'1} = \frac{0'005}{0'1} = \frac{1}{20} = 5\%$$

c)

$$P(\overline{C}|\overline{L}) = \frac{P(\overline{C} \cap \overline{L})}{P(\overline{L})} = \frac{0'405}{1 - P(L)} = \frac{0'405}{1 - 0'5} =$$

$$= \frac{0'405}{0'5} = \frac{405}{500} = \frac{81}{100} = 0'81 = 81\%$$

Ja que es té que:

$$1 - P(C) = P(\overline{C}) = P(\overline{C} \cap L) + P(\overline{C} \cap \overline{L}) = 0'495 + P(\overline{C} \cap \overline{L}) \Rightarrow$$

$$\Rightarrow p(\overline{C} \cap \overline{L}) = 0'9 - 0'495 = 0'405.$$

e) Per últim:

$$P(\overline{C}|L) = \frac{P(\overline{C} \cap L)}{P(L)} = \frac{0'495}{0'5} = \frac{495}{500} = \frac{99}{100} = 0'99 = 99\%$$

D'aquest exemple teòric es dedueix que -de no haver-hi cap llei de protecció ecològica al fet d'haver-la- la probabilitat de què l'explotació porcina no contamini el medi ambient augmenta des del 81% (apartat c)) al 99% (apartat d)), per la qual cosa es considera molt interessant la promulgació al país d'una llei d'aquestes característiques.

5.6. Cinquè problema

"Dues subparcel·les d'una finca de mandariners del Baix Ebre A i B, formades cadascuna d'elles per 100 arbres de la varietat Clemenules sobre peu o patró de *Citrange Troyer*, que ocupen aproximadament 1 jornal de terra superficial (mesura del país, 1 j_t = 2.190 m^2) cada una, pateixen una infestació severa de nematodes detectada un cop efectuats els anàlisis de terra i arrels corresponents. S'administra un nematocida a la subparcel·la A, però no a la B (que s'anomena *grup control*), essent en la resta de treballs propis del conreu (reg, aplicació de fitosanitaris, adobats, poda, recol·lecció, ...) les dues zones tractades idènticament. Es troba que en els grups A i B, 75 i 65 arbres respectivament s'han recuperat de la plaga. Assajar la hipòtesi de què el nematocida ajuda a curar la infestació al nivell de significació del (a) 0’01, (b) 0’05, (c) 0’10, (d) solucionar el mateix problema per aplicació de la prova del *txi-quadrat* i e) calcular el coeficient de contingència de Pearson.”

Solució:

Anomenarem p_1 i p_2, respectivament, les proporcions arbòries curades: (1) utilitzant el nematocida, (2) sense utilitzar el nematocida. Cal decidir, doncs, entre les dues hipòtesis següents:

> $H_0 : p_1 = p_2$, i les diferències observades són degudes al atzar, és a dir, el nematocida no és pas efectiu.
>
> $H_1 : p_1 > p_2$, i el nematocida és efectiu.

Sota la hipòtesi H_0,

$$\mu_{P1-P2} = 0 \ ; \ \ \sigma_{P1\text{-}P2} = \sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \sqrt{(0'70)(0'30)\left(\frac{1}{100}+\frac{1}{100}\right)} = 0'0648$$

on s’ha utilitzat com a estima de la proporció *p* de cures a les dues subparcel·les mostrals el valor (75 + 65)/200 = 0’70 i on q = 1 – p = 0’30. Llavors es tindrà que:

$$Z = \frac{P_1 - P_2}{\sigma_{P1-P2}} = \frac{0'75 - 0'65}{0'0648} = 1'54$$

(a) Consultades les taules corresponents de la distribució normal, i d’acord amb un assaig unilateral al nivell de significació del 0’01, es rebutjaria la hipòtesi H_0 solament si z fos més gran de 2’33. Ja que el valor de z és 1’54, es deduirà que els resultats són causa de l’atzar en aquest nivell de significació.

(b) D’acord amb un assaig unilateral al nivell de significació del 0’05, es rebutjaria la hipòtesi H_0 solament si z fos més gran de 1’645. D’on es dedueix que en aquest nivell també les diferències són causa de l’atzar.

(c) Si s’utilitzés un assaig unilateral al nivell de significació del 0’10, es rebutjaria la hipòtesi H_0 solament si el valor de z fos superior a 1’28. Donat que aquesta condició és satisfeta, es deduiria que el nematocida és efectiu al nivell de significació del 0’10.

S’ha d’advertir que les condicions anteriors depenen d’allò que s’estigui disposat a arriscar de prendre una decisió. Si els resultats venen induïts realment per l’atzar i es pren la decisió de que són deguts al nematocida (error del tipus I), es pot procedir a inocular el nematocida als arbres malgrat que resulta inefectiu. Aquest és un risc que no sempre desitgem suposar.

D'altra banda, es pot deduir que el nematocida no ajuda quan realment sí ho fa (error del tipus II). Aquesta decisió és molt important, especialment per les seves conseqüències econòmiques en la gestió de la producció de la finca.

Si ara, per tal d'assegurar les resultats, volem fer l'assaig a un major nombre d'arbres, i considerem dues subparcel·les de 300 arbres cadascuna, recuperant-se de la plaga 225 exemplars del grup A i 195 del grup B, es tindrà que, en aquest cas, la proporció d'arbres curats en els dos grups és, respectivament:

225/300 = 0'75 (grup A) i 195/300 = 0'65 (grup B)

que són les mateixes que les abans considerades. Així, doncs, sota la hipòtesi H_0, es tindrà que:

$$\mu_{P1-P2} = 0 \;;\quad \sigma_{P1\text{-}P2} = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{(0'70)(0'30)\left(\frac{1}{300} + \frac{1}{300}\right)} = 0'0374$$

on (225 – 195)/600 = 0'70 que s'empra com *estima* de *p*. Aleshores, es tindrà:

$$Z = \frac{P_1 - P_2}{\sigma_{P1-P2}} = \frac{0'75 - 0'65}{0'0374} = 2'67$$

Ara bé, tenint en compte que el valor obtingut de z és superior a 2'33, es pot rebutjar la hipòtesi al nivell de significació del 0'01, és a dir, es dedueix que el nematocida resulta efectiu amb només una probabilitat de l'1% d'error. Això posa de manifest de quina manera en incrementar-se la grandària de la mostra augmenta també la seguretat de les decisions a prendre. Tanmateix, en alguns casos ens serà força difícil augmentar aquesta grandària i llavors ens veurem obligats a prendre decisions en base a la informació disponible, tot corrent un major risc d'equivocar-nos.

(d) Les condicions del problema es presenten a la taula següent (la primera). Sota la hipòtesi nul·la H_0 de què el nematocida no té efecte, caldria esperar que 70 arbres de cadascuna de les subparcel·les és recuperessin i 30 a cada subparcel·la no es recuperessin, com s'indica a la taula corresponent (la segona). S'ha d'advertir que H_0 és equivalent a afirmar que la recuperació és independent de la utilització del nematocida, és a dir, que les classificacions són independents.

Vegem ara les esmentades taules:

FREQÜÈNCIES REALMENT OBSERVADES

	Es recuperen	No es recuperen	TOTAL
Grup A (utilitzant nematocida)	75 (a_1)	25 (a_2)	100 (n_A)
Grup B (no utilitzant nematocida)	65 (b_1)	35 (b_2)	100 (n_B)
TOTAL	140 (n_1)	60 (n_2)	200 (n)

FREQÜÈNCIES ESPERADES SOTA LA HIPÒTESI H_0

	Es recuperen	No es recuperen	TOTAL
Grup A (utilitzant nematocida)	70 (n_1n_A/n)	30 (n_2n_A/n)	100 (n_A)
Grup B (no utilitzant nematocida)	70 (n_1n_B/n)	30 (n_2n_B/n)	100 (n_B)
TOTAL	140 (n_1)	60 (n_2)	200 (n)

$$\chi^2 = \frac{(75-70)^2}{70} + \frac{(65-70)^2}{70} + \frac{(25-30)^2}{30} + \frac{(35-30)^2}{30} = 2'38$$

Per a determinar el nombre de graus de llibertat, s'ha de considerar la taula següent, que és igual a les dues donades anteriorment, però en la qual només s'han posat els totals. És clar que solament es té llibertat per a col·locar un número en una de les quatre caselles buides, puix que una vegada fet això els números de les restants caselles vénen obligats pels totals ja indicats. De tal manera que hi ha un grau de llibertat.

Així, tindrem:

	Es recuperen	No es recuperen	TOTAL
Grup A (utilitzant nematocida)			100
Grup B (no utilitzant nematocida)			100
TOTAL	140	60	200

Consultant ara la taula corresponent de la distribució de probabilitat χ^2 que trobarem a l'annex 5 del nostre llibre. Com que $\chi_{.95}$ = 3'84 per a 1 grau de llibertat i ja que χ^2 = 2'38 < 3'84, es dedueix que els resultats no són significatius al nivell del 0'05. No s'està, així, en condicions de rebutjar la hipòtesi H_0 a aquest nivell i es dedueix que o bé el nematocida no és efectiu o bé es deixa sense prendre cap decisió en espera de posteriors assajos.

S’ha de fer constar que χ^2 = 2’38 és el quadrat del valor de z = 1’54 obtingut en aquest mateix problema. En general, la prova *txi-quadrat* en relació amb dues proporcions mostrals d’una taula de contingència 2 x 2 equival a un assaig de significació de diferències de proporcions mitjançant l’aproximació normal. En efecte, es denota per P_1 i P_2 les dues proporcions mostrals i *p* la proporció poblacional.

En referència al problema que ens ocupa, es té:

(1) $$P_1 = \frac{a_1}{n_1},\ P_2 = \frac{a_2}{n_2},\ 1\text{-}P_1 = \frac{b_1}{n_1},\ 1\text{-}P_2 = \frac{b_2}{n_2}$$

(2) $$p = \frac{n_A}{n},\ 1\text{-}p = q = \frac{n_B}{n}$$

de manera que,

(3) $$a_1 = n_1 P_1,\ \ a_2 = n_2 P_2,\ \ b_1 = n_1(1-P_1),\ \ b_2 = n_2(1-P_2)$$

(4) $$n_A = n\cdot p,\quad n_B = n\cdot q$$

Utilitzant ara les expressions anteriors (3) i (4), es té el següent:

$$\chi^2 = \frac{n(a_1 b_2 - a_2 b_1)^2}{n_1 n_2 n_A n_B} = \frac{n\left[n_1 P_1 n_2 (1-P_2) - n_2 P_2 n_1 (1-P_1)\right]^2}{n_1 n_2 npnq} =$$

$$= \frac{n_1 n_2 (P_1 - P_2)^2}{npq} = \frac{(P_1 - P_2)^2}{pq(1/n_1 + 1/n_2)} \quad \text{(ja que } n = n_1 + n_2\text{)}$$

que és el quadrat de l’estadístic Z.

S’ha de fer constar també que un assaig unilateral utilitzant χ^2 equival a un assaig bilateral utilitzant χ, ja que, per exemple, $\chi^2 > \chi^2_{.95}$ correspon a $\chi > \chi_{.95}$ o $\chi < -\chi_{.95}$. Tenint en compte que per a les taules 2 x 2, χ^2 és el quadrat del valor de z, es dedueix que χ és el mateix que z en aquest cas. Així doncs, el fet de rebutjar una hipòtesi al nivell de significació del 0’05 utilitzant χ^2 equival a rebutjar aquesta hipòtesi amb un assaig unilateral al nivell de significació del 0’10 utilitzant z.

Si ara ens plantegem de solucionar el problema aplicant la correcció de Yates per a la continuïtat, ens trobarem que:

$$\chi^2(\text{corregida}) = \frac{(|75-70|-0'5)^2}{70} + \frac{(|65-70|-0'5)^2}{70} + \frac{(|25-30|-0'5)^2}{30} + \frac{(|35-30|-0'5)^2}{30} = 1'93$$

, obtenint-se que les conclusions del problema anterior són també perfectament vàlides aquí. Això s'hauria pogut veure ràpidament sense necessitat d'efectuar cap càlcul, ja que la correcció de Yates sempre disminueix -per la seva pròpia definició o naturalesa- el valor de χ^2.

Aquest mateix problema es pot resoldre per aplicació de les fórmules que es dedueixen a continuació. En principi, es pot aplicar la formulació:

$$\chi^2 = \frac{n(a_1b_2 - a_2b_1)^2}{n_1n_2n_An_B}$$

on els seus paràmetres seran els de la taula següent:

RESULTATS REALMENT OBSERVATS

	I	II	TOTAL
A	a_1	a_2	n_A
B	b_1	b_2	n_B
TOTAL	n_1	n_2	n

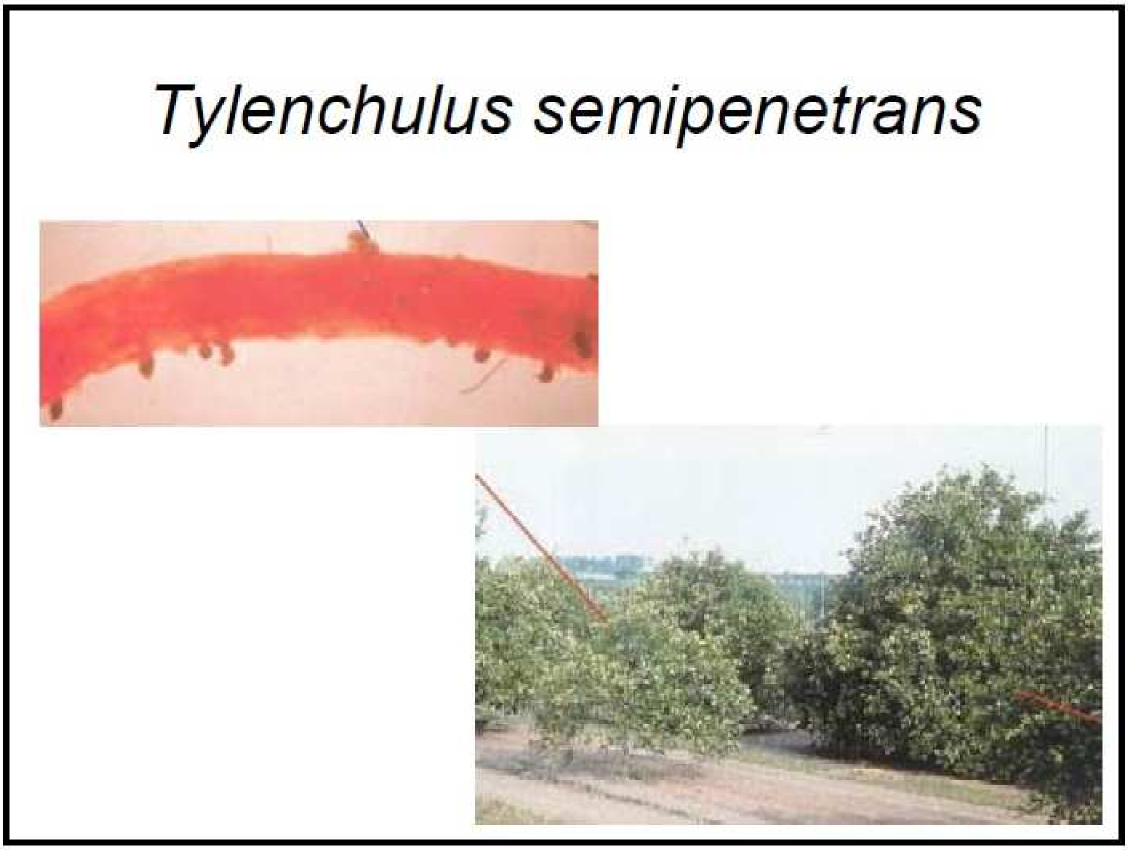


FIG. 9.25. Atac de nematodes en un camp de cítrics.

Així doncs, els resultats esperats sota la hipòtesi nul·la apareixen a la taula següent. Llavors, es tindrà que:

RESULTATS ESPERATS

	I	II	TOTAL
A	$n_1 n_A/n$	$n_2 n_A/n$	n_A
B	$n_1 n_B/n$	$n_2 n_B/n$	n_B
TOTAL	n_1	n_2	n

$$\chi^2 = \frac{(a_1 - n_1 n_A/n)^2}{n_1 n_A/n} + \frac{(a_2 - n_2 n_A/n)^2}{n_2 n_A/n} + \frac{(b_1 - n_1 n_B/n)^2}{n_1 n_B/n} + \frac{(b_2 - n_2 n_B/n)^2}{n_2 n_B/n}$$

Però: $$a_1 - \frac{n_1 n_A}{n} = a_1 - \frac{(a_1 + b_1)(a_1 + a_2)}{a_1 + b_1 + a_2 + b_2} = \frac{a_1 b_2 - a_2 b_1}{n}$$

Anàlogament: $$a_2 - \frac{n_2 n_A}{n} = b_1 - \frac{n_1 n_B}{n} = b_2 - \frac{n_2 n_B}{n} = \frac{a_1 b_2 - a_2 b_1}{n}$$

Així, es pot escriure,

$$\chi^2 = \frac{n}{n_1 n_A}\left(\frac{a_1b_2 - a_2b_1}{n}\right)^2 + \frac{n}{n_2 n_A}\left(\frac{a_1b_2 - a_2b_1}{n}\right)^2 +$$
$$+ \frac{n}{n_1 n_B}\left(\frac{a_1b_2 - a_2b_1}{n}\right)^2 + \frac{n}{n_2 n_B}\left(\frac{a_1b_2 - a_2b_1}{n}\right)^2$$

que al simplificar, dóna la formulació abans esmentada:

$$(1) \quad \chi^2 = \frac{n(a_1b_2 - a_2b_1)^2}{n_1n_2n_An_B} = \frac{n\Delta^2}{n_1n_2n_An_B}$$

on: $\Delta = a_1b_2 - a_2b_1$, $n = a_1 + a_2 + b_1 + b_2$, $n_1 = a_1 + b_1$, $n_2 = a_2 + b_2$, $n_A = a_1 + a_2$, $n_B = b_1 + b_2$. Altrament, si s'aplica la correcció de Yates, l'expressió anterior (1) es reemplaça per,

$$(2) \quad \chi^2(\text{corregida}) = \frac{n(|\Delta| - \frac{1}{2}n)^2}{n_1 \cdot n_2 \cdot n_A \cdot n_B}$$

Al problema que ens ocupa, es tenen els següents valors: $a_1 = 75$, $a_2 = 25$, $b_1 = 65$, $b_2 = 35$, $n_1 = 140$, $n_2 = 60$, $n_A = 100$, $n_B = 100$, i $n = 200$; llavors, la fórmula (1) del problema dóna el següent resultat:

$$\chi^2 = \frac{200\,[(75)(35) - (25)(65)]^2}{(140)(60)(100)(100)} = 2'38$$

Utilitzant ara la correcció de Yates, el resultat és el mateix que l'obtingut en el problema anterior, o sigui:

$$\chi^2(\text{corregida}) = \frac{n(|a_1b_2 - a_2b_1| - \frac{1}{2}n)^2}{n_1n_2n_An_B} = \frac{200[|(75)(35) - (25)(65)| - 100]^2}{(140)(60)(100)(100)} = 1'93$$

e) Sobre el significat de l'estadígraf χ^2 es pot consultar l'annex 5 d'aquest mateix llibre. Altrament, pel que es refereix al coeficient C de contingència, degut a Karl Pearson, vegem que el resultat de *txi-quadrat* té com a limitació que no permet conèixer l'adreça de l'associació. Una altra limitació d'aquest estadístic és que no permet conèixer el grau d'associació existent entre les dues variables del problema plantejat.

Una forma de conèixer el grau d'associació entre dues variables és calcular el *coeficient de contingència C.* El seu càlcul, una vegada s'ha construït una taula de contingència i calculat el valor de *txi-quadrat*, resulta molt directe, ja que n'hi ha prou amb aplicar la senzilla fórmula:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \text{ , } \forall \text{ C / } 0 \leq C < 1$$

No obstant això, les limitacions del coeficient de contingència C, que cal tenir en compte, són les següents:

1) El seu valor mínim pot ser zero, però el màxim és diferent d'1.
2) El límit superior depèn del nombre de files i columnes de la taula; així, quan el nombre de files i columnes és igual, el valor màxim és igual a l'arrel quadrada de (k-1)/k, essent k el nombre de columnes o de files. Conseqüentment, en una taula 2 x 2 l'arrel quadrada de ½ és 0,7071; en una taula de 3 x 3 és 0,8165. Els valors C de diverses taules només són comparables si les taules tenen el mateix nombre de files i columnes.
3) S'apliquen les mateixes restriccions que en el càlcul de *txi-quadrat* sobre el percentatge de cel·les amb freqüències esperades baixes.

Una forma alternativa d'aconseguir un coeficient el valor del qual se situï entre 0 i 1 és mitjançant la determinació del *coeficient V de Cramer.* Per al seu càlcul cal considerar un valor *t* que representa el valor més petit de les dues quantitats (r-1) o (s-1), on *r* i *s* són el nombre de columnes i de files. Però això, de moment, no ens ho demanen, per la qual cosa obviarem aquí la seva especificació metodològica. En el nostre cas, doncs, el coeficient de contingència de Pearson ve donat per la següent expressió:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{2'38}{2'38 + 200}} = \sqrt{0'01176} = 0'1084$$

o bé emprant l'expressió corregida de Yates:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{1'93}{1'93 + 200}} = \sqrt{0'00956} = 0'0978$$

De fet, el valor màxim de C es presenta quan les dues classificacions són perfectament dependents o associades. En tal cas, tots els arbres tractats amb nematocida es recuperaran i tots els que no es tracten no es recuperaran. Aquesta taula de contingència és la següent:

	Es recuperen	No es recuperen	TOTAL
Grup A (utilitzant nematocida)	100	0	100
Grup B (no utilitzant nematocida)	0	100	100
TOTAL	100	100	200

Com sigui que les freqüències esperades, suposant independència total, són totes iguals a 50, resultarà que:

$$\chi^2 = \frac{(100-50)^2}{50} + \frac{(0-50)^2}{50} + \frac{(0-50)^2}{50} + \frac{(100-50)^2}{50} = 200$$

Llavors el valor màxim de C és: $\sqrt{\chi^2/(\chi^2+n)} = \sqrt{200/(200+200)} = 0'7071$.

En general, per a la dependència total en una taula de contingència, en la que el número de files i de columnes són ambdues iguals a k (la nostra és 2x2), les úniques freqüències de caselles que no són cero apareixen a la diagonal (principal) que baixa d'esquerra a dreta de la taula. Per a aquests casos tenim que,

$$C_{màx} = \sqrt{(k-1)/k} = \sqrt{1/2} = 0'7071,$$

com es volia demostrar.