Powtórzenie: ANOVA 1

JEDNOCZYNNIKOWA ANALIZA WARIANCJI, ANOVA 1

Obserwowana (badana) cecha — Y

Czynnik wpływający na Y (badany) — A

A_i — **i–ty poziom** czynnika A (i=1..a),

n_i — **liczba powtórzeń** w i–tej populacji

Cecha Y ~ $N(m_i, \sigma^2)$ w **i-tej populacji** (tzn: A_1: $Y_1 \sim N(m_1, \sigma^2)$; ...; A_i: $Y_i \sim N(m_i, \sigma^2)$; ...; A_a: $Y_a \sim N(m_a, \sigma^2)$)

Model liniowy ANOVA1

$$\mathbf{y_{ij} = m + a_i + e_{ij}}$$

($\mathbf{m_i=m+a_i}$)

i = 1, 2, ..., a   **j** = 1, 2, ..., n_i

$\mathbf{y_{ij}}$ – obserwacja przeprowadzona dla i-tego poziomu czynnika A (A_i), w j – tym powtórzeniu,

m – średnia ogólna,

$\mathbf{a_i}$ – efekt i-tego poziomu czynnika A,

$\mathbf{e_{ij}}$ – efekt losowy (błąd losowy)

WIELOCZYNNIKOWA ANALIZA WARIANCJI, ANOVA

Obserwowana (badana) cecha — Y

Czynniki wpływające na Y (badane) — A, B, ...

A_i — **i–ty poziom** czynnika A (i=1,..,a)

a — liczba poziomów czynnika A

B_i — **i–ty poziom** czynnika B (i=1,..,b)

b — liczba poziomów czynnika B

...

n_{ki} — **liczba powtórzeń** wykonanych dla k–tego czynnika (A,B,…) na poziomie i–tym.

Cecha $Y \sim N(m_{ki},\sigma^2)$ dla k–tego czynnika na i–tym poziomie.

Obserwacje są niezależne w ramach powtórzeń.

Model liniowy 2–czynnikowej analizy wariancji

$$\mathbf{y_{ijk} = m + a_i + b_j + (ab)_{ij} + e_{ijk}}$$

$\mathbf{m_{ij} = m + a_i + b_j + (ab)_{ij}}$

i = 1, 2,..., a **j** = 1, 2,..., b k=1, 2,..., n_{ij}

$\mathbf{y_{ijk}}$ – obserwacja przeprowadzona dla i–tego poziomu czynnika A (A_i) i j–tego poziomu czynnika B (B_j), w k–tym powtórzeniu.

m – średnia ogólna,

$\mathbf{a_i}$ – efekt **główny** i–tego poziomu czynnika A,

$\mathbf{b_j}$ – efekt **główny** j–tego poziomu czynnika B,

$\mathbf{(ab)_{ij}}$ – efekt **interakcji** i–tego poziomu czynnika A z j–tym poziomem czynnika B

$\mathbf{e_{ijk}}$ – efekt losowy (błąd losowy)

dla jednoznaczności wyniku zakładamy

$$\sum_{i=1}^{a} a_i = 0 \quad \sum_{j=1}^{b} b_i = 0 \quad \sum_{i=1}^{a} (ab)_{ij} = 0 \quad \sum_{j=1}^{b} (ab)_{ij} = 0$$

Rozkład efektów losowych powinien być rozkładem $N(0,\sigma_{err}^{2})$ (rozkłady niezależne dla różnych poziomów czynników A i B).

Średnie brzegowe: średnia dla i–tego poziomu czynnika A: $m_{i.}=m+a_i$; średnia dla j–tego poziomu czynnika B: $m_{.j}=m+b_j$

Hipoteza 1: efekt interakcji czynnika A z czynnikiem B jest zerowy.

H_0: $(ab)_{ij}=0$ dla każdego *i* i *j*

Hipoteza alternatywna: $H_1 : \exists_{\, i,j} \;\; (ab)ij \neq 0$

Hipoteza 2: efekty główne czynnika A są zerowe (czynnik A nie wpływa na średnie poprzez poziomy czynnika B wartości cechy Y)

$H_0 : m_{1.} = m_{2.} = \ldots = m_{a.} = m$

Hipoteza alternatywna: $H_1 : \exists_{\, i'} \;\; m_{i'.} \neq m$

Hipoteza 3: efekty główne czynnika B są zerowe

$H_0 : m_{.1} = m_{.2} = \ldots = m_{.b} = m$

Hipoteza alternatywna: $H_1 : \exists_{\, j'} \;\; m_{.j'} \neq m$

Jeżeli hipoteza o efektach interakcji zostanie odrzucona, wtedy oba czynniki mają wpływ na cechę Y. Jeśli nie będzie odrzucona (nieistotna interakcja) to możliwy jest jeszcze wpływ efektów głównych każdego z czynników (hipoteza 2 i 3).

Funkcje testowe dla badanej hipotezy

Funkcje testowe opierają się na hipotezie o równości wariancji spowodowanej danym czynnikiem (efekt interakcji czynnika A z czynnikiem B, efekt główny czynnika A, efekt główny czynnika B) i wariancji losowej (wywołanej odpowiednim błędem losowym).

W zależności od **planu doświadczenia** (modelu statystycznego tego doświadczenia) odpowiednio określa się wariancję dla błędów losowych służących do testowania poszczególnych hipotez (o istotności odpowiednich efektów).

Każdy czynnik może być traktowany jako czynnik stały (ang. *fixed*), tzn. jego poziomy są w jakiś sposób kontrolowane i mają ustalony wpływ, albo czynnik losowy (ang. *random*), którego efekty uznajemy za losowe (tzn. o średnich zgodnych z rozkładem N).

Czynnik **stały**: Z reguły liczba poziomów czynnika stałego jest niewielka. W badaniu uwzględniamy z góry określone poziomy czynnika i wnioski odnosimy wyłącznie do tych poziomów czynnika. Przykładem czynnika stałego może być: płeć, rasa.

Czynnik **losowy**: Liczba poziomów czynnika losowego jest zwykle duża. Badaniom poddany jest losowy podzbiór wszystkich poziomów czynnika. Wnioski odnosimy do wszystkich potencjalnie możliwych poziomów. Przykładem są warunki sezonu wegetacyjnego wpływające na jakość płodów rolnych.

Czynniki mogą być niezależne (doświadczenie w układzie całkowicie losowym krzyżowym) lub jeden z czynników może być zmieniany (zagnieżdżony) w ramach drugiego czynnika (głównego, nadrzędnego) – hierarchiczna ANOVA.

Tabela analizy wariancji ANOVA dla doświadczenia w układzie krzyżowym całkowicie losowym z równą liczbą powtórzeń (n) w podgrupach

Źródło zmienności *source*	**Sumy kwadratów** *SS*	**stopnie swobody** *df*	**średni kwadrat** *MS*	$\mathbf{F_{emp}}$
Czynnik A	SS_A	$V_A=a-1$	$s_A^2 = \frac{SS_A}{V_A}$	$\frac{s_A^2}{s_e^2}$
Czynnik B	SS_B	$V_B=b-1$	$s_B^2 = \frac{SS_B}{V_B}$	$\frac{s_B^2}{s_e^2}$
Interakcja AB	SS_{AB}	$V_{AB}=(a-1)*(b-1)$	$s_{AB}^2 = \frac{SS_{AB}}{V_{AB}}$	$\frac{s_{AB}^2}{s_e^2}$
Błąd losowy	SS_e	$V_e=ab(n-1)$	$s_e^2 = \frac{SS_e}{V_e}$	
Zmienność całkowita	SS_y	$V_T=abn-1$		

Jako poziom istotności a wybiera się najczęściej wartości: 0,05 i 0,01 (oznaczając poszczególne F_{emp} jako * lub **). Często dodaje się kolumnę z wartościami statystyki P (prawdopodobieństwami popełnienia błędu I rodzaju, P–value).

P-Value

Do tabeli analizy wariancji dodaje się kolumnę z wyliczoną wartością tej statystyki gdyż wnioskowanie często odbywa się na podstawie wartości statystyki określającej **prawdopodo-bieństwo** (ryzyko) **popełnienia błędu I rodzaju** przy aktualnie obserwowanych danych. Jest to (def.) prawdopodobieństwo, że dana statystyka będzie miała conajmniej taką jak obserwowana wartość przy założeniu, że H_0 jest prawdziwa.

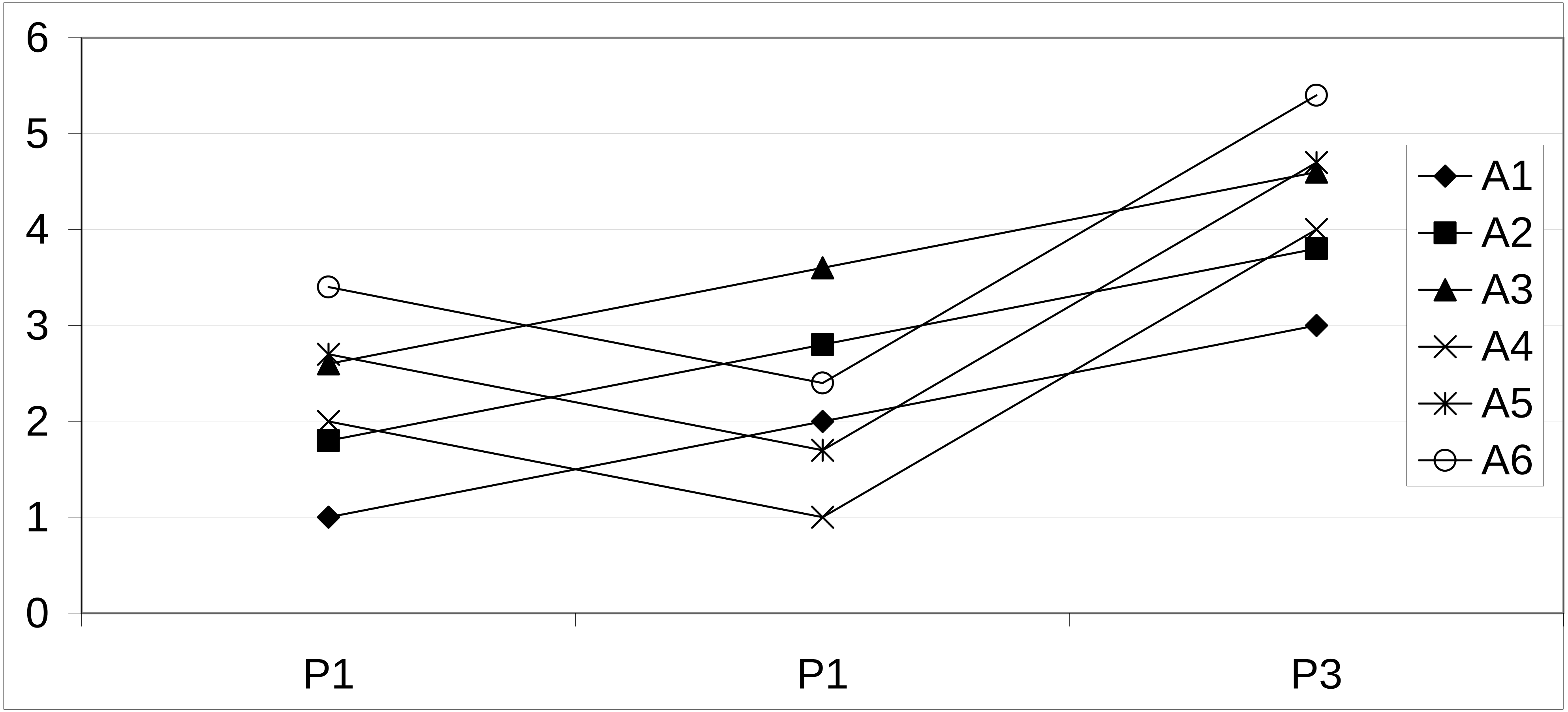
Podobnie jak w 1–czynnikowej analizie wariancji możliwe jest zastosowanie testu porównań wielokrotnych dla wybranego (istotnego) czynnika. Jednak w przypadku istotności efektów interakcji grupy o jednorodnych średnich brzegowych nie stanowią pełnego opisu zjawiska, więc postępowanie takie ma **ograniczone** zastosowanie.

Przykład: Badano zgodność 3 metod określenia łącznej zawartości tłuszczu w mleku UHT. Wybrano losowo po 12 kartonów mleka tłustego, średniotłustego i chudego. Zastosowano 3 metodu pomiaru (na 4 losowych kartonach każdego rodzaju mleka). Analizę chemiczną wykonywało 4 laborantów.

Analysis of Variance for TŁUSZCZ - Type I Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:METODA	**0,11945**	**2**	**0,059725**	**3,87**	**0,0350**
B:MLEKO	**45,9829**	**2**	**22,9914**	**1488,45**	**0,0000**
C:LABORANT	**0,0400083**	**3**	**0,0133361**	**0,86**	**0,4736**
INTERACTIONS					
AB	**0,18745**	**4**	**0,0468625**	**3,03**	**0,0371**
RESIDUAL	**0,370717**	**24**	**0,0154465**		
TOTAL	**46,7005**	**35**			

Poglądowy wykres średnich podklas dla danych zawierających efekty interakcji



Układ 1–czynnikowy całkowicie losowy

czynnik na 4 poziomach (A,B,C i D)
Liczba powtórzeń n=4

A1	B1	C1	A2
D1	A3	D2	C2
B2	D3	C3	B3
C4	A4	B4	D4

16 razy robimy roztwór (w każdej probówce osobno), i do każdej probówki dodajemy (częściowo losowo) wybrany środek chemiczny A,B,C lub D, po 4 razy każdy)

Tabela ANOVA

Source of variation	**Degrees of freedom**[a]	**Sums of squares (SSQ)**	**Mean square (MS)**	**F**
Treatments (*Tr*)	t-1	SS_{Tr}	$MS_{Tr}= SS_{Tr}/(t-1)$	MS_{Tr}/MS_E
Error (*E*)	t*(r-1)	SS_E	$MS_E= SS_E/(t*(r-1))$	
Total (*Tot*)	t*r-1	SS_{Tot}		

[a]where t=number of treatments and r=number of replications per treatment.

Układ 1–czynnikowy losowanych bloków

czynnik na 4 poziomach (A,B,C i D)
Liczba powtórzeń n=4

```
Blok I       A    B    C    D
Blok II      D    A    B    C
Blok III     B    D    C    A
Blok IV      C    A    B    D
```

4 razy robimy roztwór (w dużych probówkach), i z każdej z nich nalewamy ten prawie identyczny roztwór do 4 małych probówek. Do każdej małej probówki z 4 w bloku dodajemy inny środek chemiczny A,B,C lub D (losowo)

Tabela ANOVA

Source of variation	**Degrees of freedom**[a]	**Sums of squares (SSQ)**	**Mean square (MS)**	**F**
Blocks (*B*)	b-1	SS_B	$MS_B = SS_B/(b-1)$	MS_B/MS_E
Treatments (*Tr*)	t-1	SS_{Tr}	$MS_{Tr} = SS_{Tr}/(t-1)$	MS_{Tr}/MS_E
Error (*E*)	(t-1)*(b-1)	SS_E	$MS_E = SS_E/((t-1)*(b-1))$	
Total (*Tot*)	t*b-1	SS_{Tot}		

[a]where t=number of treatments and b=number of blocks or replications.

układ 2-czynnikowy całkowicie losowy

1. czynnik na 2 poziomach (A,B)
2. czynnik na 2 poziomach (a,b)

Liczba powtórzeń n=4

Aa1	Ba1	Ab1	Aa2
Bb1	Aa3	Bb2	Ab2
Ba2	Bb3	Ab3	Ba3
Ab4	Aa4	Ba4	Bb4

16 razy robimy roztwór (w każdej probówce osobno), i do każdej probówki dodajemy wybrany kwas A lub B oraz wybraną zasadę a lub b (losowa kolejność).

Tabela ANOVA

Source of variation	Degrees of freedom[a]	Sums of squares (SSQ)	Mean square (MS)	F
First factor (*A*)	a-1	SS_A	$MS_A = SS_A/(a-1)$	MS_A/MS_E
Second factor (*B*)	b-1	SS_B	$MS_B = SS_B/(b-1)$	MS_B/MS_E
First X Second (*AxB*)	(a-1)*(b-1)	SS_{AxB}	$MS_{AxB} = SS_{AxB}/(a-1)*(b-1)$	MS_{AxB}/MS_E
Error (*E*)	a*b*(n-1)	SS_E	$MS_E = SS_E/(a*b*(n-1))$	
Total (*Tot*)	a*b*n-1	SS_{Tot}		

[a]where a=number of treatments in the first factor, b=number of treatments in the second factor and n=number of replications.

układ 2-czynnikowy losowanych bloków

1. czynnik na 2 poziomach (A,B)
2. czynnik na 2 poziomach (a,b)

Liczba bloków n=4

Block I	Aa	Ba	Ab	Bb
Block II	Bb	Aa	Ba	Ab
Block III	Ba	Bb	Ab	Aa
Block IV	Ab	Aa	Ba	Bb

4 razy robimy roztwór (w dużych probówkach), i z każdej z nich nalewamy ten prawie identyczny roztwór do 4 małych probówek. Do każdej małej probówki z 4 w bloku dodajemy wybrany kwas, A lub B, oraz wybraną zasadę, a lub b, w losowej kolejności.

Tabela ANOVA

Source of variation	**Degrees of freedom**[a]	**Sums of squares (SSQ)**	**Mean square (MS)**	**F**
Blocks (*B*)	b-1	SS_B	$MS_B = SS_B/(b-1)$	MS_B/MS_E
First factor (*F1*)	f-1	SS_{F1}	$MS_{F1}=SS_{F1}/(f-1)$	MS_{F1}/MS_E
Second factor (*F2*)	s-1	SS_{F2}	$MS_{F2}=SS_{F2}/(s-1)$	MS_{F2}/MS_E
First X Second (*FxS*)	(f-1)(s-1)	SS_{FxS}	$MS_{FxS}= SS_{FxS}/((f-1)*(s-1))$	MS_{FxS}/MS_E
Error (*E*)	(fs-1)(b-1)	SS_E	$MS_E= SS_E/((f*s-1)*(b-1))$	
Total (*Tot*)	fsb-1	SS_{Tot}		

[a]where f=number of treatments in the first factor, s=number of treatments in the second factor and b=number of blocks or replications.

układ 2-czynnikowy split-plot

1. czynnik na 3 poziomach (A,B,C)
2. czynnik na 2 poziomach (a,b)

Liczba bloków n=4

blok I	blok II	blok III	blok IV
Aa Ab	Bb Ba	Ab Aa	Ca Cb
Ca Cb	Ab Aa	Ba Bb	Cb Ca
Bb Ba	Aa Ab	Ca Cb	Ba Bb

4 razy robimy roztwór (w dużych probówkach), i z każdej z nich nalewamy ten prawie identyczny roztwór do 3 średnich probówek. Do każdej z nich (losowo) dodajemy wybrany kwas, A,B lub C. Każdą z tych średnich probówek rozlewamy do 2 małych. Do każdej z takich 2 probówek wlewamy (losowo) inną zasadę (a lub b)

Tabela ANOVA

Source of variation	**Degrees of freedom**[a]	**Sums of squares (SSQ)**	**Mean square (MS)**	**F**
bloki	n-1	SS_R		
czynnik A	a-1	SS_A	$MS_A = SS_A/(a-1)$	MS_A / MS_{E1}
błąd I	(n-1)(a-1)	SS_{E1}	$MS_{E1} = SS_{E1}/(n-1)(a-1)$	
czynnik B	b-1	SS_B	$MS_B = SS_B/(b-1)$	MS_B / MS_{E2}
czynnik A x czynnik B (AxB)	(a-1)(b-1)	SS_{AxB}	$MS_{AxB} = SS_{AxB} / (a-1)*(b-1)$	MS_{AxB} / MS_{E2}
błąd II	a(n-1)(b-1)	SS_{E2}	$MS_{E2} = SS_{E2} / a*(n-1)*(b-1)$	
Razem (*Tot*)	abn-1	SS_{Tot}		
Gdzie n to liczba bloków, a liczba poziomów czynnika A (głównego), b liczba poziomów czynnika B (rozlosowanego w podblokach czynnika A)				

układ 2-czynnikowy split-block

1. czynnik na 3 poziomach (A,B,C)
2. czynnik na 2 poziomach (1,2)

oba czynniki zagnieżdżone w blokach
Liczba powtórzeń n=3

Blok I		Blok II		Blok III	
A1	A2	C2	C1	B1	B2
B1	B2	A2	A1	C1	C2
C1	C2	B2	B1	A1	A2

Tabela ANOVA

Source of variation	Degrees of freedom[a]	Sums of squares (SSQ)	Mean square (MS)	F
Bloki (*B*)	b-1	SS_B	$MS_B = SS_B/(b-1)$	MS_B/MS_{AxB}
Czynnik 1 (*A*)	a-1	SS_A	$MS_A = SS_A/(t-1)$	MS_A/MS_{AxB}
Czynnik 1x Bloki (*AxB*)	(a-1)*(r-1)	SS_{AxB}	$MS_{AxB} = SS_{AxB}/((t-1)*(r-1))$	
Czynnik 2 (*C*)	c-1	SS_C	$MS_C = SS_C/(c-1)$	MS_C/MS_{CxB}
Czynnik 2 X Bloki (*CxB*)	(c-1)*(r-1)	SS_{CxB}	$MS_{CxB} = SS_{CxB}/((c-1)*(r-1))$	
Czynnik 1 X Czynnik 2 (*AxC*)	(a-1)*(c-1)	SS_{AxC}	$MS_{AxC} = SS_{AxC}/((c-1)*(t-1))$	MS_{AxC}/MS_E
Błąd (*E*)	(a-1)*(c-1)*(b-1)	SS_E	$MS_E = SS_E/((t-1)*(c-1)*(b-1))$	
Razem (*Tot*)	a*c*b-1	SS_{Tot}		

[a] gdzie a= liczba poziomów czynnika A, c= liczba poziomów czynnika B, b= liczba bloków.

Tabela analizy wariancji dla modelu ANOVA 2–czynnikowego o losowanych blokach. Bloki są zagnieżdżone w czynniku L. Modele 1-4 związane są z uznaniem czynników (G lub L) jako czynniki stałe lub losowe

ANOVA models including the factors G = genotype and L = location or environment, and estimation of variance components, for trials in a randomized complete block design

Source	DF[e]	MS	Model 1[a]		Model 2[b]		Model 3[c]		Model 4[d]
			F test	Variance component	F test	Variance component	F test	Variance component	F test
G	g-1	M1	M1/M4	sg2 = (M1-M4)/rl	M1/M4	-	M1/M5	sg2 = (M1-M5)/rl	M1/M5
L	l-1	M2	M2/M4	-	M2/M5	-	M2/M4	-	M2/M5
Block (L)	(r-1)*l	M3	-	-	-	-	-	-	-
G × L	(g-1)*(l-1)	M4	M4/M5	sgl2 = (M4-M5)/r	M4/M5	sgl2 = (M4-M5)/r	M4/M5	sgl2 = (M4-M5)/r	M4/M5
Pooled error	(r-1)*(g-1)*l	M5	-	se2 = M5	-	se2 = M5	-	se2 = M5	-

[a] Model 1 = G and L random factors; [b] Model 2 = G fixed, L random; [c] Model 3 = L fixed, G random; [d] Model 4 = G and L fixed factors. [e] *g* = no. genotypes; *l* = no. locations; *r* = no. blocks.