第七章

基因型和环境互作

基 本 概 念

- **环境 (Environment):** 环境定义为影响一个基因型表现的一组非遗传因素。
  - 非生物因素, 如土壤的物理和化学特性、气候因子（如光照，降雨量和温度）等
  - 生物因素, 包含害虫、病原体、线虫和杂草等
- 这样定义的环境又称宏环境（Macro-environment）

基 本 概 念

- 微环境 (Micro-environment): 微环境定义为单个植株或小区的生长环境，两个不同的植株或小区生长在同样的微环境中生长的可能性几乎是0

基 本 概 念

- 但一般来说，宏环境间的差异要比微环境间的差异大很多，基因型和环境的互作研究中的环境一般指的是宏环境，微环境的效应一般视为随机误差效应。
- 宏环境可以是不同的栽培方式、地点和年份，也可以是不同的栽培方式、地点和年份的组合。

基本概念

- **可预测的环境变异**
  - 可预测的环境变异包括一些永久性质的环境因素，如气候的周期性变化、土壤类型、日照时间等，还有一些栽培措施也可看作可预测的环境变异，如播种时间、播种密度、施肥水平、收获方式等；
- **不可预测的环境变异**
  - 不可预测的环境变异包括一些气候的变动因素如降雨量和温度以及病虫害的侵袭等。

基 本 概 念

- 目标环境群体 (TPE， target population of environments)
  - 品种种植环境的总称
  - TPE可能是由多种环境类型构成，每种环境类型内部差异相对较小的（或称为同质的），但是不同环境类型间可能有较大差异
  - 基因的效应在一种环境类型内差异不大，但在不同环境类型间可能有较大差异

基因型与环境交互作用的模型

- 表型值分解

$$y_{ijk} = \mu + g_i + e_j + (ge)_{ij} + \varepsilon_{ijk}$$

- 互作效应估计

$$(ge)_{ij} = y_{ij\bullet} - g_i - e_j - \mu$$

- 观测值的方差分解

$$V_y = V_G + V_E + V_{GE} + V_\varepsilon$$

基因型效应和环境效应的估计

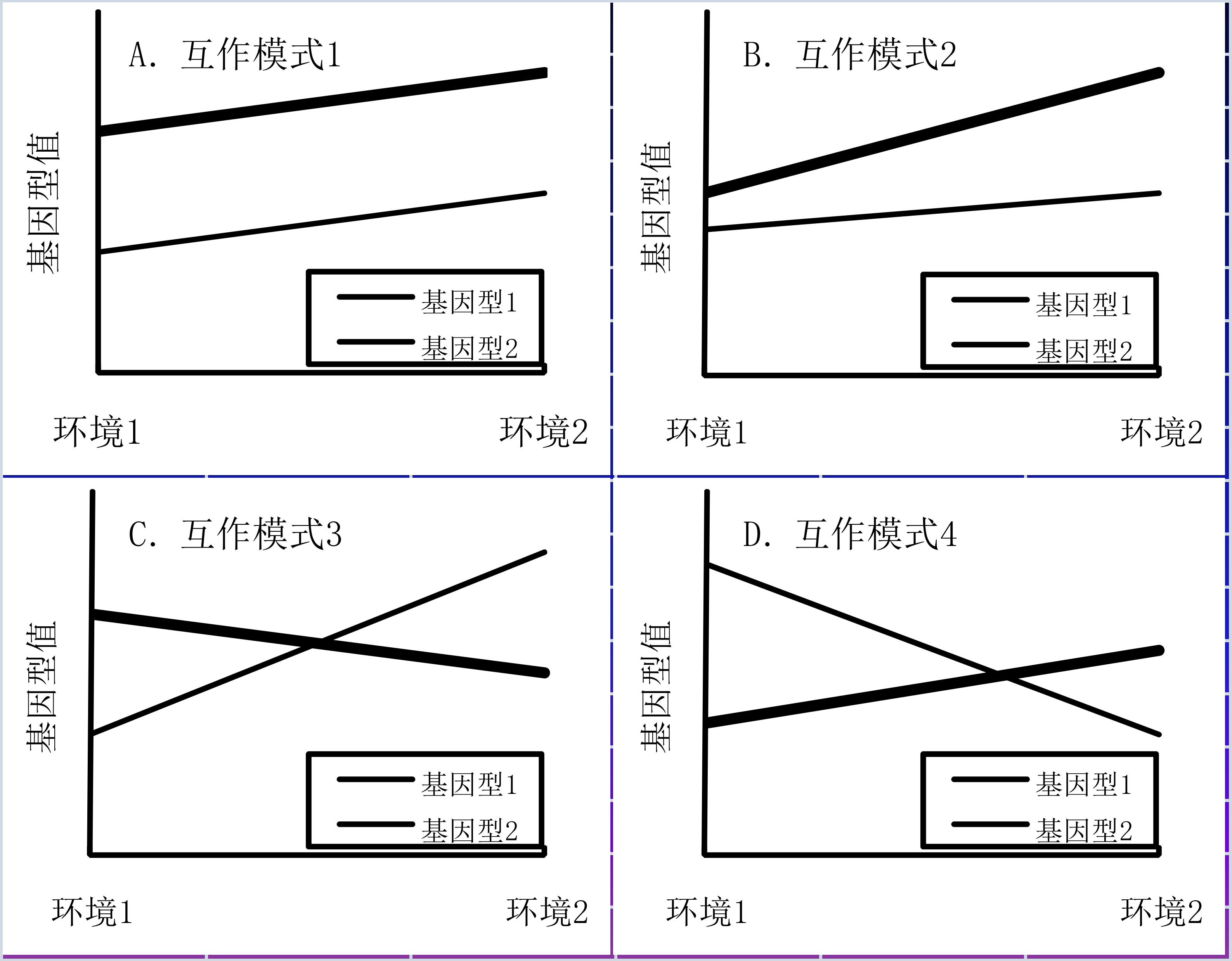
	环境 1	环境 2	平均	基因型效应（g_i）
基因型 1	$y_{11\bullet}$	$y_{12\bullet}$	$y_{1\bullet\bullet}$	$g_1 = y_{1\bullet\bullet} - \mu$
基因型 2	$y_{21\bullet}$	$y_{22\bullet}$	$y_{2\bullet\bullet}$	$g_2 = y_{2\bullet\bullet} - \mu$
基因型 3	$y_{31\bullet}$	$y_{32\bullet}$	$y_{3\bullet\bullet}$	$g_3 = y_{3\bullet\bullet} - \mu$
平均	$y_{\bullet1\bullet}$	$y_{\bullet2\bullet}$	$\mu = y_{\bullet\bullet\bullet}$	
环境效应（e_j）	$e_1 = y_{\bullet1\bullet} - \mu$	$e_2 = y_{\bullet2\bullet} - \mu$		

两种小麦基因型在2个环境小麦赤霉病感染率（%）

	环境1	环境2	平均	基因型效应
基因型1	10	30	20	g1=-15
基因型2	50	50	50	g2=15
平均	30	40	m=35	0
环境效应	e1=-5	e2=5	0	

- 交互作用的估计
  - ✓ $(ge)_{11}=y_{11}-g_1-e_1-m=10-(-15)-(-5)-35=-5$
  - ✓ $(ge)_{12}=y_{12}-g_1-e_2-m=30-(-15)-5-35=5$
  - ✓ $(ge)_{21}=y_{21}-g_2-e_1-m=50-15-(-5)-35=5$
  - ✓ $(ge)_{22}=y_{22}-g_2-e_2-m=50-15-5-35=-5$

基因型与环境互作的四种模式



利用基因型与环境互作的途径

- 基因型与环境交互作用是广泛存在的，实际中有3种方式来应对基因型与环境交互作用
  - 方式1：忽略基因型与环境互作 (Ignore it)
  - 方式2：降低基因型与环境互作 (Reduce it)
  - 方式3：利用基因型与环境互作 (Exploit it)

基因型和环境试验的方差分析

变异来源	自由度	均方	随机模型期望均方
环境间（E）	$e-1$		
环境内重复间（R/E）	$e(r-1)$		
基因型间（G）	$n-1$	MS_G	$V_\varepsilon + rV_{GE} + reV_G$
基因型×环境互作（GE）	$(n-1)(e-1)$	MS_{GE}	$V_\varepsilon + rV_{GE}$
随机误差	$(n-1)(r-1)e$	MS_ε	V_ε

评价基因型的环境数和重复数的确定

- 基因型和环境的互作方差

$$V_{GE} = \frac{MS_{GE} - MS_{\varepsilon}}{r}$$

- 最小显著性差异 $LSD_{\alpha} = t_{\alpha/2}[df = (n-1)(e-1)]\sqrt{\frac{2MS_{GE}}{re}}$

$$LSD_{0.05} \approx 2.0\sqrt{2(\frac{V_{\varepsilon}}{re} + \frac{V_{GE}}{e})}$$

- 基因型的个数较多时，

$$e = \frac{8}{(LSD_{0.05})^2}(\frac{V_{\varepsilon}}{r} + V_{GE})$$

t=2.0，e 的大小为

评价基因型的环境数和重复数的确定

- 基因型平均数的方差为 $V_{\bar{y}} = \frac{V_{\varepsilon}}{re} + \frac{V_{GE}}{e}$
- 最小显著差异 $LSD_{0.05} \approx 2.0\sqrt{2V_{\bar{y}}}$
  - 在总的小区数*re* 固定的情况下，最优的资源配置是令重复数 *r* =1，而使环境数e尽可能的大

目标环境群体(TPE)的分类

- 分类分析 (Classification procedure)
- 主坐标分析 (Ordination procedure)
- 联合分类和坐标过程 (Joint use of a classification procedure and an ordination procedure)

两个环境间的距离

$$D_{jj'} = \tfrac{1}{n}\sum_{i=1}^{n}\left(\frac{y_{ij}-\mu_j}{SE_j}-\frac{y_{ij'}-\mu_{j'}}{SE_{j'}}\right)^2$$

$$D_{jj'} = 2\left(1-\tfrac{1}{n}\right)\left(1-r_{jj'}\right)$$

以四个环境为例说明环境的分组

4个环境的距离矩阵

	环境2	环境3	环境4
环境1	0.88	1.20	1.20
环境2		1.18	1.28
环境3			1.33

- **类平均法 (Average linkage，也称UPGMA)**
  - **步骤1：把两个最相似的环境合并成第一个类即环境1和环境2。**

步骤2：把两个最相似的类合并

- **3种可能的合并是：**
  - (环境1—环境2)—环境3，平均距离为(D13+D23)/2=(1.20+1.18)/2=1.19；
  - (环境1—环境2)—环境4，平均距离为(D14+D24)/2=(1.20+1.28)/2=1.24；
  - 环境3—环境4，平均距离为D34=1.33。

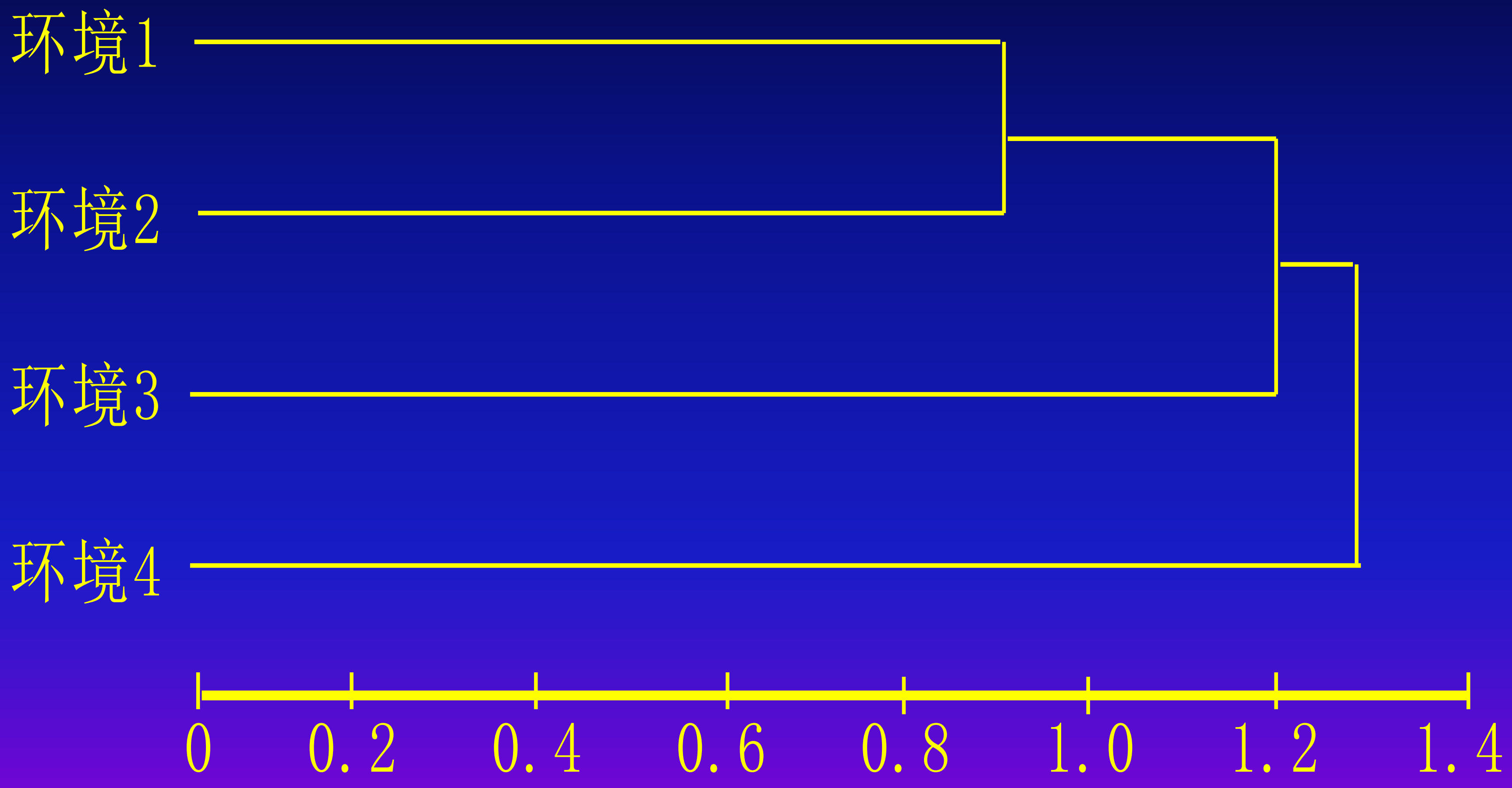
- 结果: (环境1—环境2)—环境3

步骤3：把4个环境合并为一个类

- 合并成1个大类后的类平均距离为：

$$(D14+D24+D34)/3=1.27$$

四个环境的聚类图



基因型和环境互作的联合回归分析

- **环境指数 (Environment index) 的定义**

	环境 1	环境 2	…	环境 e	平均数	基因型效应
基因型 1	y_{11}	y_{12}	…	y_{1e}	$y_{1\bullet}$	$g_1 = y_{1\bullet} - \mu$
基因型 2	y_{21}	y_{22}	…	y_{2e}	$y_{2\bullet}$	$g_2 = y_{2\bullet} - \mu$
…	…	…	…	…	…	…
基因型 n	y_{n1}	y_{n2}	…	y_{ne}	$y_{n\bullet}$	$g_n = y_{n\bullet} - \mu$
平均数	$y_{\bullet 1}$	$y_{\bullet 2}$	…	$y_{\bullet e}$	$\mu = y_{\bullet\bullet}$	
环境效应	$e_1 = y_{\bullet 1} - \mu$	$e_2 = y_{\bullet 2} - \mu$		$e_e = y_{\bullet e} - \mu$		

Finlay和Wilkinson模型

- 回归系数的估计为 $$b_i=\frac{\sum_j(\bar{y}_j-\bar{y})(y_{ij}-\bar{y}_i)}{\sum_j(\bar{y}_j-\bar{y})^2}=\frac{\sum_j y_{ij}\bar{y}_j-\frac{1}{e}(\sum_j y_{ij})(\sum_j \bar{y}_j)}{\sum_j \bar{y}_j^2-\frac{1}{e}(\sum_j \bar{y}_j)^2}$$
- Finlay和Wilkinson提供的衡量基因型随环境变化的灵敏性的方法：
  - 当*b*=1表示品种对环境的反应等于所有基因型 的平均反应；
  - 当*b*＞1表示品种对环境的反应高于所有基因型的平均反应，稳定性差；
  - 当*b*＜1表示品种对环境的反应低于所有基因型的平均反应，稳定性好。

品种稳定性分析的Eberhart 和 Russell方法

- 品种稳定性分析 (Stability analysis)
- 反应参数 (Parameter of response)
- 稳定性参数 (Parameter of stability)

品种稳定性分析的Eberhart 和Russell方法

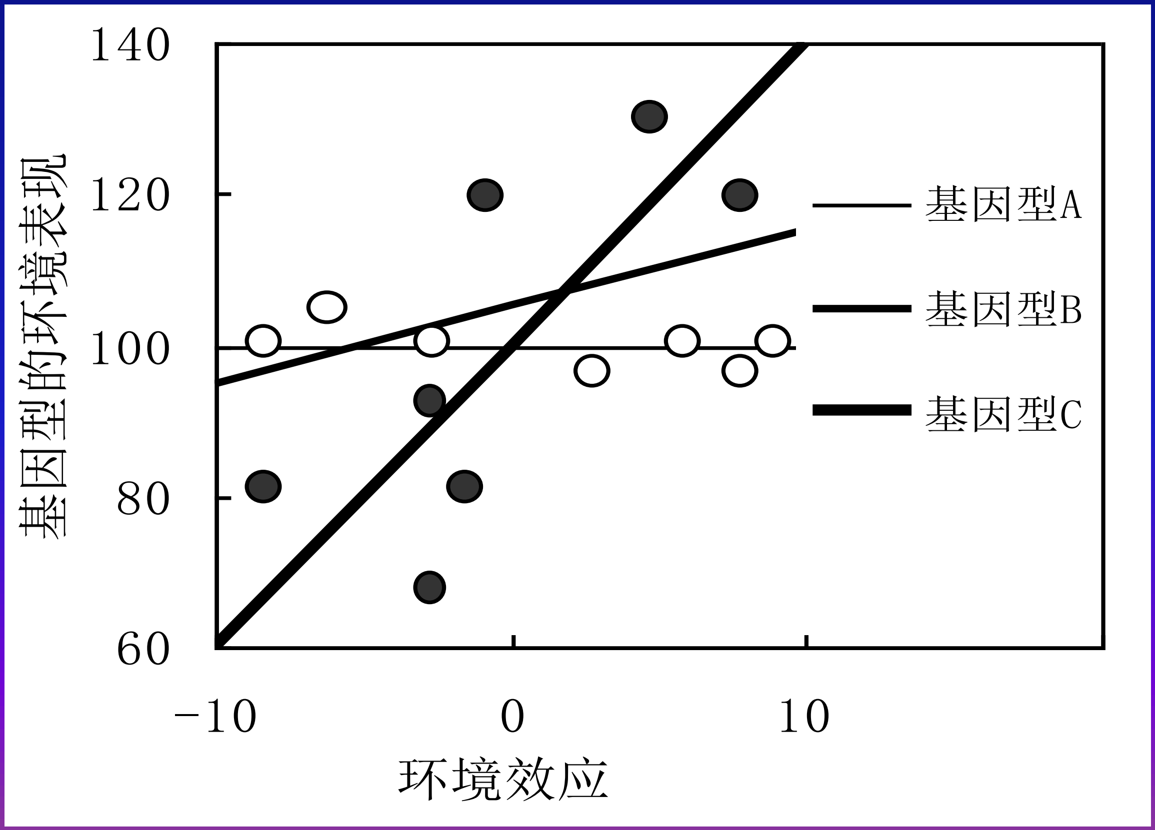
- 线性模型为 $y_{ij} = \mu_i + b_i e_j + \delta_{ij} = \mu + g_i + b_i e_j + \delta_{ij}$
- 稳定参数估计为 $b_i = \dfrac{\sum_j y_{ij} e_j}{\Sigma e_j^2}$
- 基因型*i*的表现预测为 $\hat{y}_{ij} = \mu_i + b_i e_j$
- 稳定性测度 $s^2_{\delta(i)} = \dfrac{1}{e-2} \sum_j \hat{\delta}^2_{ij} - \dfrac{1}{r} MS_\varepsilon$

联合回归分析中方差的分解

变异来源	自由度	平方和	随机模型期望均方
环境间（E）	$e-1$	$SS_E = rn\Sigma e_j^2$	$V_\varepsilon + rV_{GE} + reV_G$
基因型间（G）	$g-1$	$SS_G = re\Sigma g_i^2$	$V_\varepsilon + rV_{GE} + reV_G$
基因型×环境（GE）	$(g-1)(e-1)$	$SS_G = r\Sigma(ge)_{ij}^2$	$V_\varepsilon + rV_{GE}$
环境（线性）	1	$SS_{E(\mathrm{Re}g)}$	
品种回归	$n-1$	SS_{Reg}	
离回归	$(n-1)(e-2)$	SS_{Residual}	
随机误差	$ne(r-1)$	MS_ε	V_ε

Eberhart 和Russell的稳定性分类

- 回归系数 $b_i = 0$ 时的稳定性称为类型A
- 回归系数 $b_i = 1$时的稳定性称为类型B
- 较小的 $s_{\delta(i)}^2$ 引起的稳定性称为类型C



10个黄花烟草品系对环境指数的联合回归分析

基因型	1	2	3	4	5	6	7	8	9	10
平均表现	84.87	105.18	101.86	106.25	114.05	124.39	128.97	133.17	130.62	144.22
回归平方和	763.90	575.55	514.82	1001.31	922.87	227.33	1640.17	833.05	1022.84	1008.83
离回归平方和	108.09	61.97	404.52	213.54	41.44	71.37	144.03	149.54	101.35	174.89
回归系数	0.97	0.84	0.80	1.11	1.07	0.53	1.42	1.01	1.12	1.12
决定系数（R^2）	0.88	0.90	0.56	0.82	0.96	0.76	0.92	0.85	0.91	0.85

基因型和环境互作的乘积模型

- 主效相加互作相乘模型 (AMMI，additive main effects and multiplicative interaction)
- AMMI模型为

$$y_{ijk} = \mu + g_i + e_j + \sum_{m=1}^{N} (IPCA_m^{Gi})(IPCA_m^{Ej}) + \delta_{ij} + \varepsilon_{ijk}$$

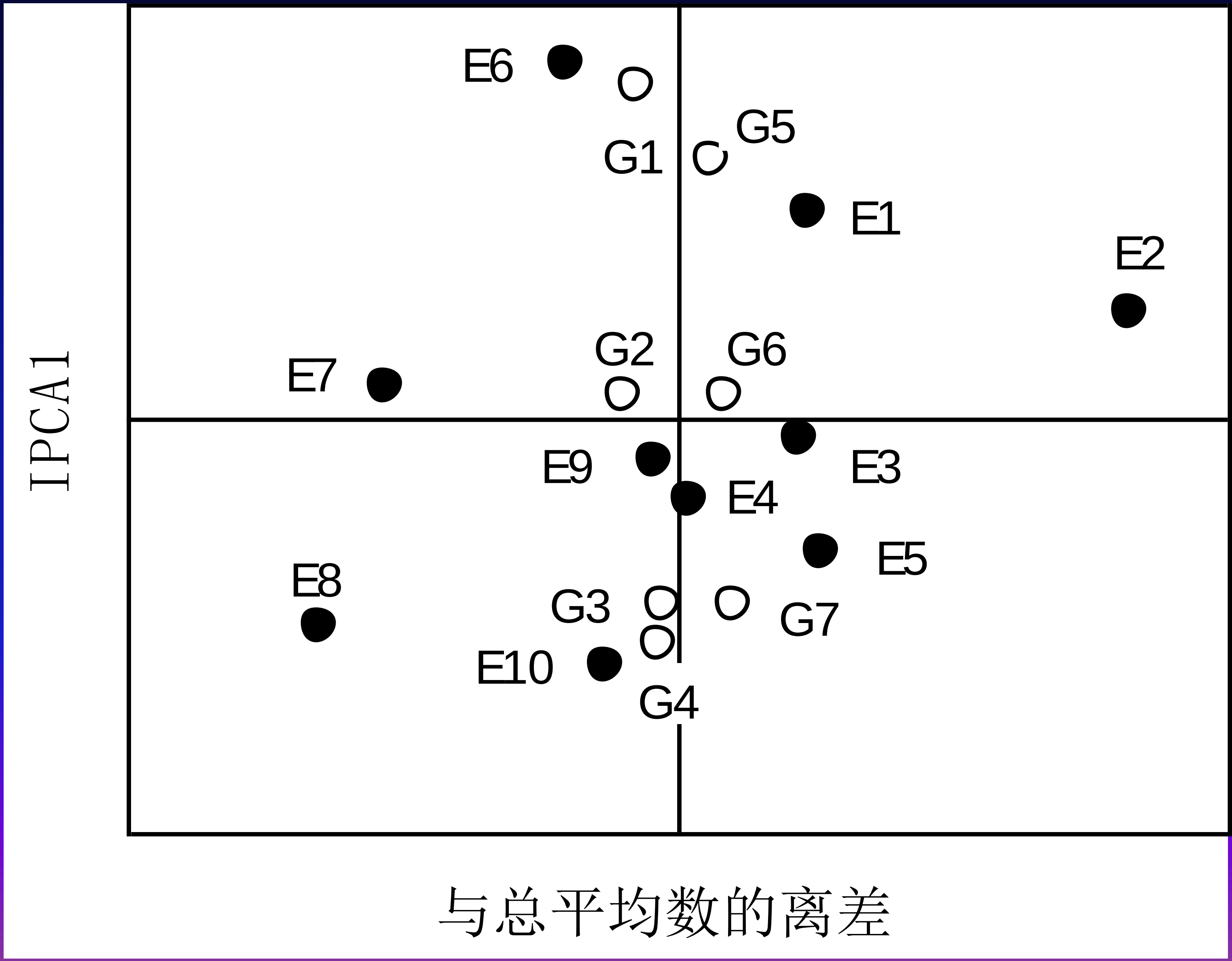
- 基因型和环境的互作效应为

$$(ge)_{ij} = \sum_{m=1}^{N} (IPCA_m^{Gi})(IPCA_m^{Ej}) + \delta_{ij}$$

- 基因型 i 在环境 j 下表型的估计值为

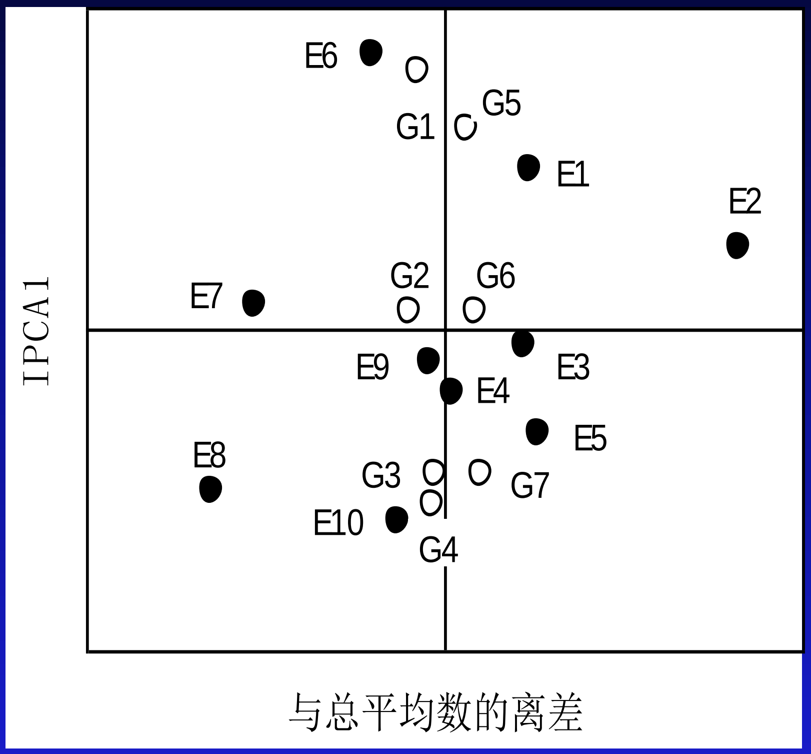
$$\hat{P}_{ij} = \mu + g_i + e_j + \sum_{m=1}^{N} (IPCA_m^{Gi})(IPCA_m^{Ej})$$

AMMI分析的双标图



7个基因型在10个环境下的双标图

- 具有较高表现的基因型和环境出现在双标图的右上方，如图中基因型5和6、环境1和2
- 如果某个基因型和某个环境的IPCA1得分同正或者同负，这时基因型和环境互作是有利的，例如基因型3和环境8和10之间的互作是有利的
- 如果某个基因型和某个环境的IPCA1得分一正一负，这时基因型和环境互作是不利的，例如基因型3和环境6和7之间的互作是不利的。



AMMI分析的有效性和存在的问题

- AMMI分析中，对主成分的数目没有固定的规则，一些经验表明*N*=1，往往就是很好的模型。
- AMMI的有用性体现在IPCA得分所能解释的可重复互作变异的大小，在有些试验中，第一主成分可以解释60-80%的互作变异，利用AMMI分析得到很好的结果；在有些试验中，第一主成分只可以解释10-30%的互作变异，利用AMMI分析不一定会得到很好的结果。

AMMI分析的有效性和存在的问题

- AMMI预测结果也依赖于互作变异中可重复部分的大小，以及可重复变异的可重复性。
  - 在前面的例子中，基因型3和环境8之间的互作是有利的，利用这一信息去预测基因型在环境8中的表现的前提是，这种有利互作在年份间是可重复的。有些环境因素年份间保持相对稳定，如土壤特性和耕作方式等，而有些环境因素，如降水量和温度等，不容易预测。因此AMMI分析利用的是年份间能够重复的互作。

GGE互作分析的乘积模型

- **GGE互作分析模型**

$$y_{ijk} = \mu + g_i + e_j + (ge)_{ij} + \varepsilon_{ijk}$$

$$g_i + (ge)_{ij} = \sum_{m=1}^{N} (IPCA_m^{Gi})(IPCA_m^{Ej}) + \delta_{ij}$$