Training course-cum-workshop on

Statistical, Biometrical and Genomic Methods

Mexico city, 22 May, 2019

Alvarado Beltran Gregorio: g.alvarado@cigar.org
Outline

- Genotype by Interaction (GE)
  - Fixed models
  - Random models
- Remarks
- References
\[ Y_{ijk} = \mu + \text{Rep}_j + \text{Block}_k(\text{Rep}_j) + \text{Gen}_i + \varepsilon_{ijk} \]

**Lattice design one environment**

\[ Y_{ijkl} = \mu + \text{Env}_l + \text{Rep}_j(\text{Env}_l) + \text{Block}_k(\text{Rep}_j \text{ Env}_l) + \text{Gen}_i + \text{Env}_l \times \text{Gen}_i + \varepsilon_{ijkl} \]

**Lattice design multi-environment**

Where;

- \( Y_{ijkl} \) is the trait of interest
- \( \mu \) is the overall mean
- \( \text{Rep}_j \) is the effect of the \( j^{th} \) replicate
- \( \text{Gen}_i \) is the effect of the \( i^{th} \) genotype; fixed-BLUEs, random-BLUPs
- \( \text{Cov} \) is the effect of the covariate, \( k^{th} \) is the block effect,
- \( \text{Env}_l \) is the site effect it could be fixed or random
- \( \varepsilon_{ijkl} \) is the effect of the error associated with all sources of variation.
• There are two components of the model

  - Additive \((\mu + Env_l + Rep_j(Env) + Block_k(Rep_j Env_l) + Gen_i)\)

  - Multiplicative (the interaction \(Env_l \times Gen_i\))
Soil
- Water
- N
- P
- K
- Micronutrients

Weather
- Temperature
- Precipitation
- Sun radiation

Plant
- Phenotype
- QTLs
- Genes
- DNA
- Biochemical pathways

Trait measured in the last stage of the crop

Source: Dr. Jose Crossa
• Phenotype is the result of specific genotype responses, to specific environmental effects.

• In multi-environment trials (METs), separability of the interaction GxE effects does not occur (i.e., there is significant COI).

• A joint multiplicative operator of the $a(\alpha)$ (genotypes) $b(\gamma)$ (environments) type needs to be defined for studying the non-separability (COI) of effects.
Therefore, the model

\[ Y_{ijkl} = \mu + Env + Rep_j(Env_l) + Block_k(Rep_j Env_l) + Gen_i + Env_l \times Gen_i + \varepsilon_{ijkl}. \]

• **Is Unparsimonious:** each GE cell has its own interaction parameter

• **Is Uninformative:** the independent interaction parameters are difficult to interpret.
• Analysis of METs could be done through using fixed, random or mixed, linear models,

\[ Y_{ijkl} = \mu + \text{Env}_l + \text{Rep}_j(\text{Env}_l) + \text{Block}_k(\text{Rep}_j \text{ Env}_l) + \text{Gen}_i + \text{Env}_l \times \text{Gen}_i + \epsilon_{ijkl}. \]

• Fixed: Ordinary Least Squares (OLS),

• Random: Maximum Likelihood (ML), REML (Restricted Maximum Likelihood).
• Some authors consider additive terms as Linear and Multiplicative terms as Bi-Linear.

• Classes of fixed effects linear-bilinear models (Crossa and Cornelius)
  
  • Additive Main Effect and Multiplicative Interaction (AMMI)
    \[
    \bar{y}_{ij} = \mu + \tau_i + \delta_j + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\epsilon}_{ij}
    \]

  • Genotypes Regression Model (GREG)
    \[
    \bar{y}_{ij} = \mu_i + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\epsilon}_{ij},
    \]
• Sites Regression Model (SREG)

\[
\bar{y}_{ij} = \mu_j + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\epsilon}_{ij}
\]

• Completely Multiplicative Model (COMM)

\[
\bar{y}_{ij} = \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\epsilon}_{ij}
\]

• Shifted Multiplicative Model (SHMM)

\[
\bar{y}_{ij} = \beta + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\epsilon}_{ij}
\]
The SHMM model is used for identifying subsets of genotypes or environments with negligible COI.

The SREG model has been used in preference to SHMM for grouping environments without genotypic rank change.

\[ \bar{y}_{ij} = \mu_j + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \varepsilon_{ij}, \]

\[ \bar{y}_{ij} = \mu_j + \lambda_1 \alpha_{1j} \gamma_{j1} + \lambda_2 \alpha_{2j} \gamma_{j2} + \lambda_3 \alpha_{3j} \gamma_{j3} + ... + \varepsilon_{ij} \]

Variability due to COI (rank change)  
Variability due to non-COI (not rank change)
FACTORIAL REGRESSION

For \( g = 1, \ldots, G \) genotypic covariables, \( X_{i1}, \ldots, X_{iG} \)

\[
\bar{y}_{ij} = \mu + G_i + E_j + \sum_{g=1}^{G} X_{ig} \xi_{jg} + \varepsilon_{ij}
\]

For \( h = 1, \ldots, H \) environmental covariables, \( z_{j1}, \ldots, z_{jH} \)

\[
\bar{y}_{ij} = \mu + G_i + E_j + \sum_{h=1}^{H} \zeta_{ih} z_{jh} + \varepsilon_{ij}
\]
For both genotypic and environmental covariables, simultaneously

\[ y_{ij} = \mu + G_i + E_j + \sum_{g=1}^{G} x_{ig} \xi_{jg} + \sum_{h=1}^{H} \zeta_{ih} Z_{jh} + \sum_{g=1}^{G} \sum_{h=1}^{H} v_{gh} x_{ig} Z_{jh} + \epsilon_{ij} \]

- Hypotheses related to the significance of the effects for the available external covariables can be tested.
- Standard variable subset selection procedures, like stepwise regression can be used for model construction.
• When environmental (genotypic) covariables show high collinearity, interpretation of the least squares regression coefficients is very imprecise.

• When the number of covariables is larger than the number of observations, least squares estimation of the parameters in the FR models is not unique.

• Partial Least Squares (PLS) regression overcomes some of these problems and it can be used as an alternative estimation method.
Bilinear Models: PLS

• When genotypic responses over environments (Y) are modeled using environmental covariables, then the matrix Z of environmental covariables can be written in a bilinear form as

\[ Z = t_1 p'_1 + t_2 p'_2 + \ldots + t_M p'_M + E_M = TP' + E \]  \hfill (6)

• The matrix T contains the \( t_1 \ldots t_j \) vectors called Z-scores (indexed by environments).

• The matrix P has the \( p_1 \ldots p_H \) vectors called Z-loadings (indexed by environmental variables) and E has the residuals.
• Similarly, the response variable matrix $Y$ in bilinear form is

$$Y = t_1 q_1' + t_2 q_2' + \ldots + t_M q_M' + F_M = TQ' + F$$  \hspace{0.5cm} (7)$$

• The matrix $Q$ contains the $q_1 \ldots q_I$ vectors called $Y$-loadings (indexed by genotypes) and $F$ has the residuals.

• The relationship between $Y$ and $Z$ is transmitted through the latent variable $T$. 
• Given that $T = ZW$; where $W$ is a vector of weights

\[
E(Y') = (TQ')' = QT' = Q(ZW)' = (QW')Z' = \zeta Z'
\]
• as the last term in (4)

• Similarly, for the $X$ matrix of genotypic covariables, $T = XW$, and

\[
E(Y) = TQ' = XWQ' = X\Xi'
\]
• as the last term in (2)
• The set of matrices $T$, $W$ and $Q$ can be depicted in the same biplot.

  • The rows of matrix $T$ contain the coordinates for environments
  • The rows of $W$ contain the coordinates for environmental covariables
  • The rows of $Q$ contain the coordinates for cultivars
Software

• To analyze METs
  • META-R (META-R for Spatial analysis)
  • SAS (META-SAS)
  • R

• Software to analyze GE
  • GEA (fixed GE)
  • ASREML (random GE)
  • SAS (random and fixed GE)
References

- Biometrics and Statistics Unit, Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), México.
• References

Thank you for your interest!