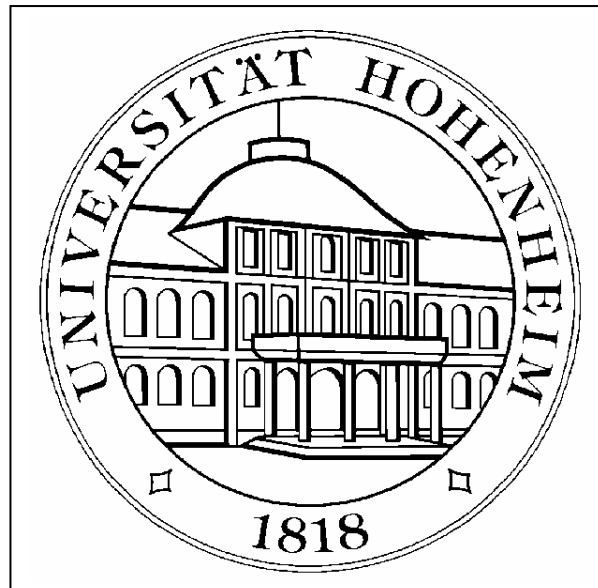


Plant breeding trials

Hans-Peter Piepho

Biostatistics Unit
Universität Hohenheim, Germany



Introduction



Phenotype \Leftrightarrow Genotype

$P = G + E$

“We should never forget who is the elephant in the room”

(Alan Schulman at kick-off meeting of COST StatSeq)

Introduction

The phenotype

Crop variety trials and plant breeding trials:

- Test performance for **target region**
- Trials in **large number of environments** (ideally random sample from target)



Standard trial designs for large number of treatments:

- Lattice designs, α -designs, row-column designs (Williams and John, 1995)
- Designs with spatial analysis in mind (Cullis et al., 2006; Williams et al., 2006)
- Unreplicated designs with checks, p-rep designs, augmented p-rep designs (Cullis et al., 2006; Williams et al., 2011, 2013)

Introduction



Estimation of genetic values

- Classical plant breeding based on phenotypic data alone (field trials)
- Hunting for single genes:
 - ⇒ Use of marker data for mapping of quantitative trait loci (QTL) in simple segregating populations, linkage mapping
 - ⇒ Association mapping in larger populations with diverse structure (multiple crosses, diverse breeding material, gene bank data)
 - ⇒ Marker-assisted selection (MAS) based on detected QTL
- Giving up the hunt:
 - ⇒ Just try to improve estimate of genotypic value (breeding value) using all (or most of the) markers

Introduction

Key idea of genomic selection (GS)



Predict genotypic value g_i of i -th genotype by regression on marker types

$$g_i = \sum_{k=1}^M u_k z_{ik} \quad (i = 1, 2, \dots, G)$$

z_{ik} = regressor variable for the i -th genotype and k -th marker ($k = 1, \dots, M$)

u_k = regression coefficients

Example: Biallelic marker (SNP) with alleles A_1 and A_2 , DH lines:

$$z_{ik} = 1 \quad \text{for } A_1A_1$$

$$z_{ik} = -1 \quad \text{for } A_2A_2$$

$$z_{ik} = 0 \quad \text{for } A_1A_2 \quad \text{or} \quad \text{when the marker genotype is missing}$$



At the risk of boring you, because this involves no marker or sequence data, I want to talk about the

Elephant in the room

first!

Overview

Three parts of lecture

- I. A pragmatic approach to formulating linear mixed models for randomized experiments
- II. Single-stage versus stage-wise analysis
- III. Estimating heritability and predictive accuracy of genomic prediction in plant breeding programs