Selection of genotypes for resistance and tolerance to pathogens:

A bivariate statistical analysis of yield and disease response.

Alison Kelly Bethany Macdonald Susan Fletcher

December 2, 2015

GRDC Grains Research & Development Corporation
Your GRDC working with you

Queensland Government

A bivariate statistical analysis of yield and disease

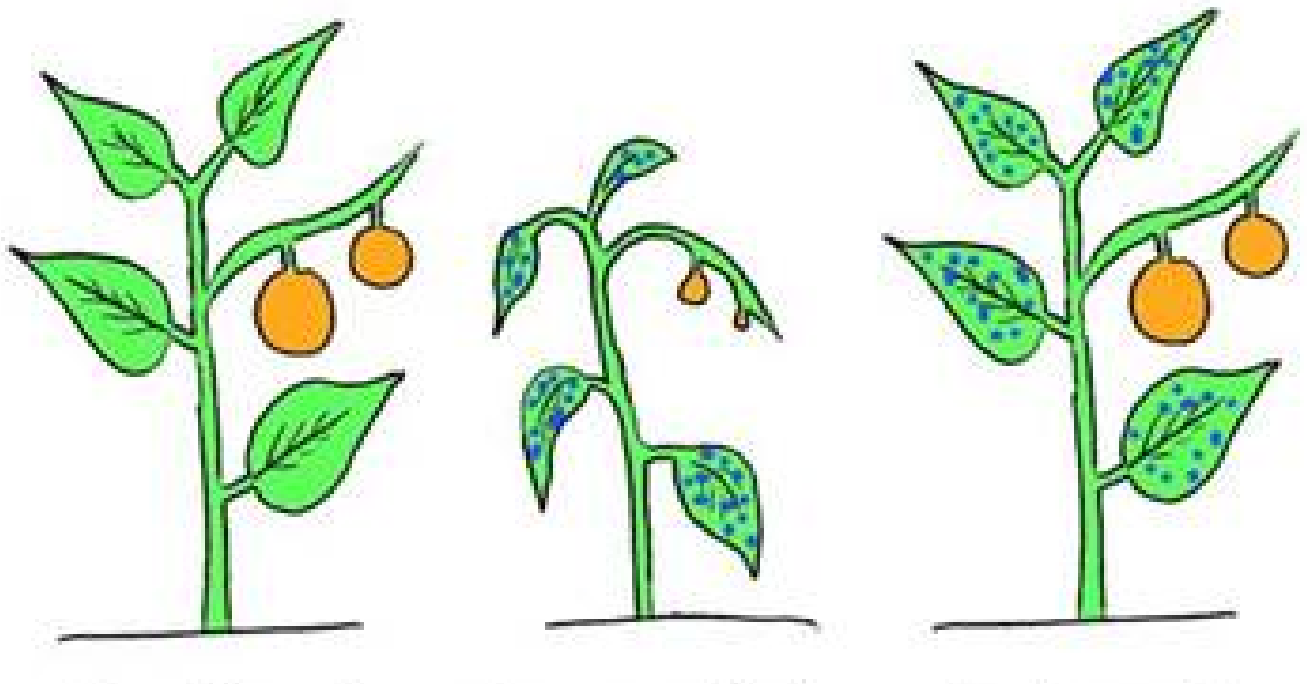
Structure

- Background to the agricultural research problem
- Linear mixed model formulation
- Interpretation of results

A bivariate statistical analysis of yield and disease
Background

Phenotyping for the effect of disease on genotypes in a plant breeding program requires measurement of

- growth of the pathogen in the plant
- subsequent effect of the pathogen on grain production in the plant



A bivariate statistical analysis of yield and disease

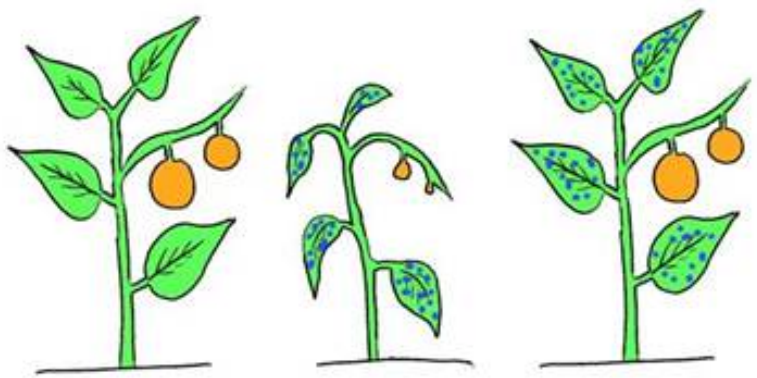
Definitions

Resistance

- The ability of the host to prevent pathogen entry and control pathogen spread.
- In this application, the ability of the plant to reduce pathogen burden and hence suppress disease expression.

Tolerance

- The ability of the host to limit the impact of pathogen burden on host health.
- In this application, the ability of the plant to produce grain in the presence of disease.



A bivariate statistical analysis of yield and disease

The statistical challenge

Uncoupling resistance and tolerance is a challenge, but is desireable for a number of reasons (Kause and Odegard, 2012).

- Each have a different impact on the relationship between the host and the pathogen.
- In plants (and animals) they are weakly genetically correlated, so are dfferent traits.
- Plant breeders should exploit both traits to provide a responsible solution to sustainable farming practices and increased food production.

Host: Wheat Pathogen: *Fusarium pseudograminearum*

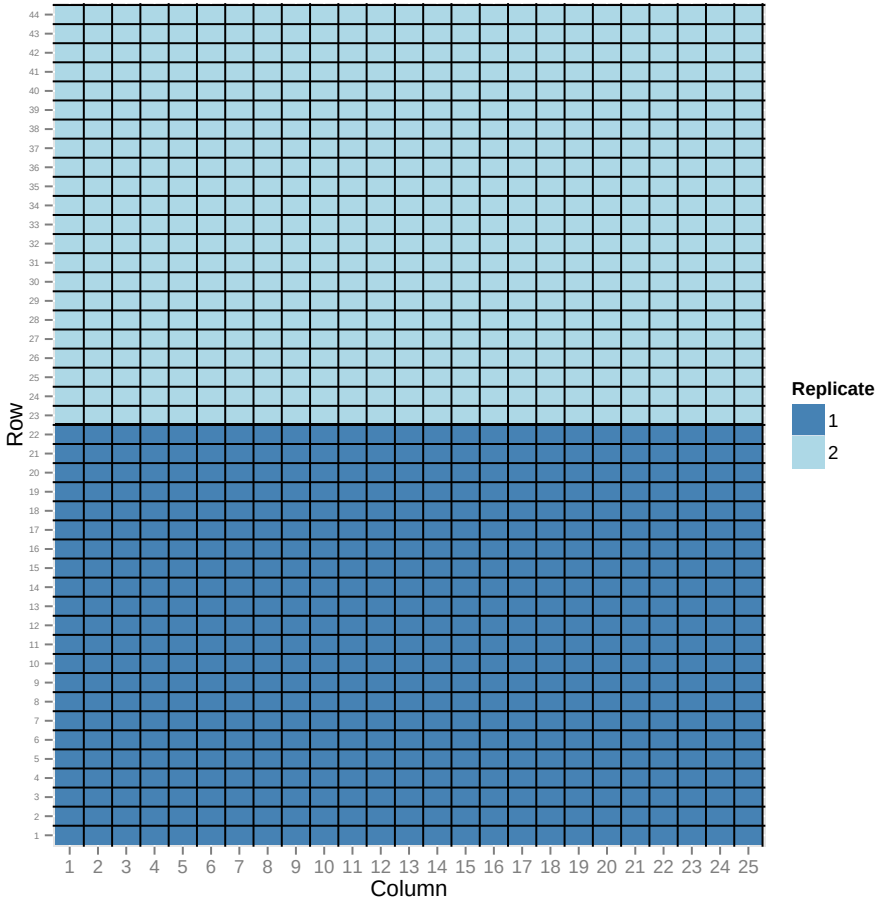
A bivariate statistical analysis of yield and disease
Experimental design

- Experiments to test the resistance and tolerance of genotypes involve a disease treatment with 2 levels: an untreated control (nil) and an imposed disease level (plus).
- The experimental design typically consists of replicated field trials in a strip-plot arrangement of the disease treatment, where genotypes are grown under these two conditions in paired plots.

- **Aim:** To select genotypes possessing combined traits of resistance and tolerance to disease.

A bivariate statistical analysis of yield and disease
Experimental design

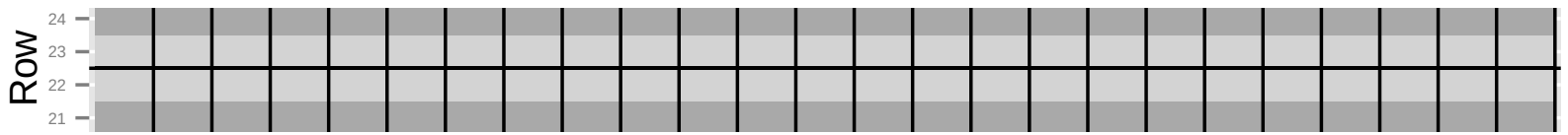
Field trial planted as a rectangular array of plots in 25 columns by 44 rows, grown at Narrabri, NSW in 2014



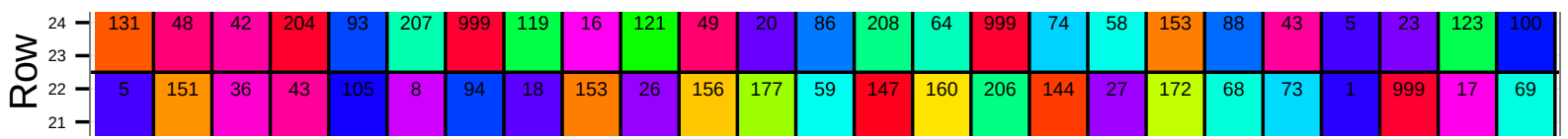
A bivariate statistical analysis of yield and disease
Experimental design

Treatment allocation for a subset of plots: 25 columns by 4 rows

Inoculum strips (nil,plus)



Genotype mainplots



A bivariate statistical analysis of yield and disease
Experimental design

Measurements

- Yield of grain (t/ha).
- Disease - level of stem browning (cm).

Issues surrounding the methodology

- A baseline treatment of nil disease aims to estimate yield potential.
- Achieving this baseline is difficult in practice due to background levels of disease in the field.
- The severity of disease expression will change with environmental conditions.

The linear mixed model
Univariate - Yield, Disease

The form of the linear mixed model for a single trait (j=Y,D) is

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\tau}_j + \mathbf{Z}_{g_j}\mathbf{u}_{g_j} + \mathbf{Z}_{s_j}\mathbf{u}_{s_j} + \mathbf{e}_j$$

$\mathbf{y}_j^{(n\times 1)}$ is the vecter of responses measured on n plots,
$\boldsymbol{\tau}_j^{(t\times 1)}$ is a vector of fixed effects for inoculum with design matrix $\mathbf{X}_j^{(n\times t)}$,
$\mathbf{u}_{g_j}^{(tm\times 1)}$ is a vector of random effects for genotype, partitioned for inoculum level, with design matrix $\mathbf{Z}_j^{(n\times m)}$,
$\mathbf{u}_{s_j}^{(s\times 1)}$ is a vector of random effects for structural terms, with design matrix $\mathbf{Z}_{s_j}^{(n\times s)}$, and
$\mathbf{u}_{s_j} = (\mathbf{u}'_{r_j}, \mathbf{u}'_{b_j}, \mathbf{u}'_{ip_j}, \mathbf{u}'_{gp_j})'$ and
$\mathbf{e}_j^{(n\times 1)}$ is the vector of residual errors.

The linear mixed model
Variance models

The variance of the random inoculum by genotype effects, $\mathbf{u}_{g_j}$, is

$$\text{var}(\mathbf{u}_{g_j}) = \begin{bmatrix} \sigma^2_{g_{n_j}} & \sigma_{g_{np_j}} \\ \sigma_{g_{np_j}} & \sigma^2_{g_{p_j}} \end{bmatrix} \otimes \mathbf{I}_m$$

The random effects for structural terms and residual errors are assumed to be independent and identically distributed normal variates with mean zero and variance structures, $\text{var}(\mathbf{u}_{s_{j_i}}) = \sigma^2_{s_{j_i}} \mathbf{I}_{n_{s_i}}$ and $\text{var}(\mathbf{e}) = \sigma^2_j \mathbf{I}_n$ (where n_{s_i} is the length of $\mathbf{u}_{s_{j_i}}$).

The linear mixed model
Bivariate - Yield and Disease

The form of the linear mixed model for both traits (j=Y,D) is

$$\mathbf{y} = (\mathbf{I}_2 \otimes \mathbf{X})\boldsymbol{\tau} + (\mathbf{I}_2 \otimes \mathbf{Z}_g)\mathbf{u}_g + (\mathbf{I}_2 \otimes \mathbf{Z}_s)\mathbf{u}_s + \mathbf{e}$$

where

$\mathbf{y} = (\mathbf{y}'_Y, \mathbf{y}'_D)'$, $\boldsymbol{\tau} = (\boldsymbol{\tau}'_Y, \boldsymbol{\tau}'_D)'$, $\mathbf{u}_g = (\mathbf{u}'_{g_Y}, \mathbf{u}'_{g_D})'$, $\mathbf{u}_s = (\mathbf{u}'_{s_Y}, \mathbf{u}'_{s_D})'$ and $\mathbf{e} = (\mathbf{e}'_Y, \mathbf{e}'_D)'$.

The bivariate linear mixed model
Variance models

The variance of the random inoculum by genotype effects, $\mathbf{u_g}$, for both traits(Y,D) is

$$\text{var}(\mathbf{u}_g) = \begin{bmatrix} \sigma^2_{g_{Y_n}} & & & \\ \sigma_{g_{Y_{np}}} & \sigma^2_{g_{Y_p}} & & \\ \sigma_{g_{Y_n D_n}} & \sigma_{g_{Y_p D_n}} & \sigma^2_{g_{D_n}} & \\ \sigma_{g_{Y_n D_p}} & \sigma_{g_{Y_p} D_p} & \sigma_{g_{D_{np}}} & \sigma^2_{g_{D_p}} \end{bmatrix} \otimes \mathbf{I}_m$$

Additionally, the variance of the random structural effects and the residual errors, $\mathbf{e}$, for both traits(Y,D) are

$$\text{var}(\mathbf{u}_{s_i}) = \begin{bmatrix} \sigma^2_{s_{Y_i}} & \\ \sigma_{s_{YD_i}} & \sigma^2_{s_{D_i}} \end{bmatrix} \otimes \mathbf{I}_{n_{s_i}}$$

$$\text{var}(\mathbf{e}) = \begin{bmatrix} \sigma^2_Y & \\ \sigma_{YD} & \sigma^2_D \end{bmatrix} \otimes \mathbf{I}_n$$

The bivariate linear mixed model

We fit the linear mixed model in ASReml-R (Butler *et al.*, 2013).

- Estimate the variance components using Residual maximum likelihood (REML) (Patterson and Thompson, 1977).
- Produce Best Linear Unbiassed Predictors of the random genotype by inoculum effects.

- Use the inherent regression structure in the bivariate analysis to interpret the results, relating this back to tolerance and resistance in the plant-pathogen context.

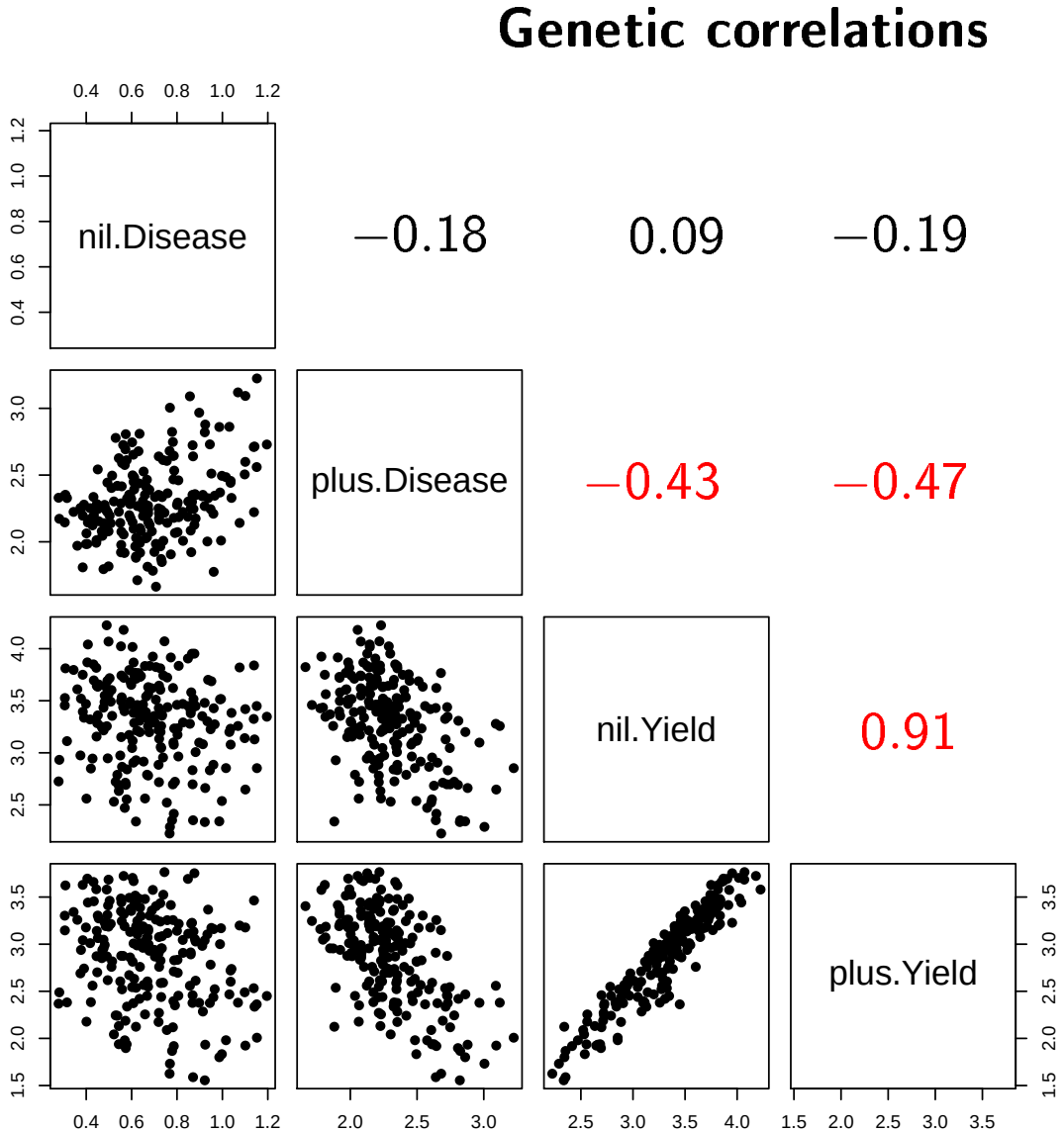
The bivariate linear mixed model

Estimates from the model

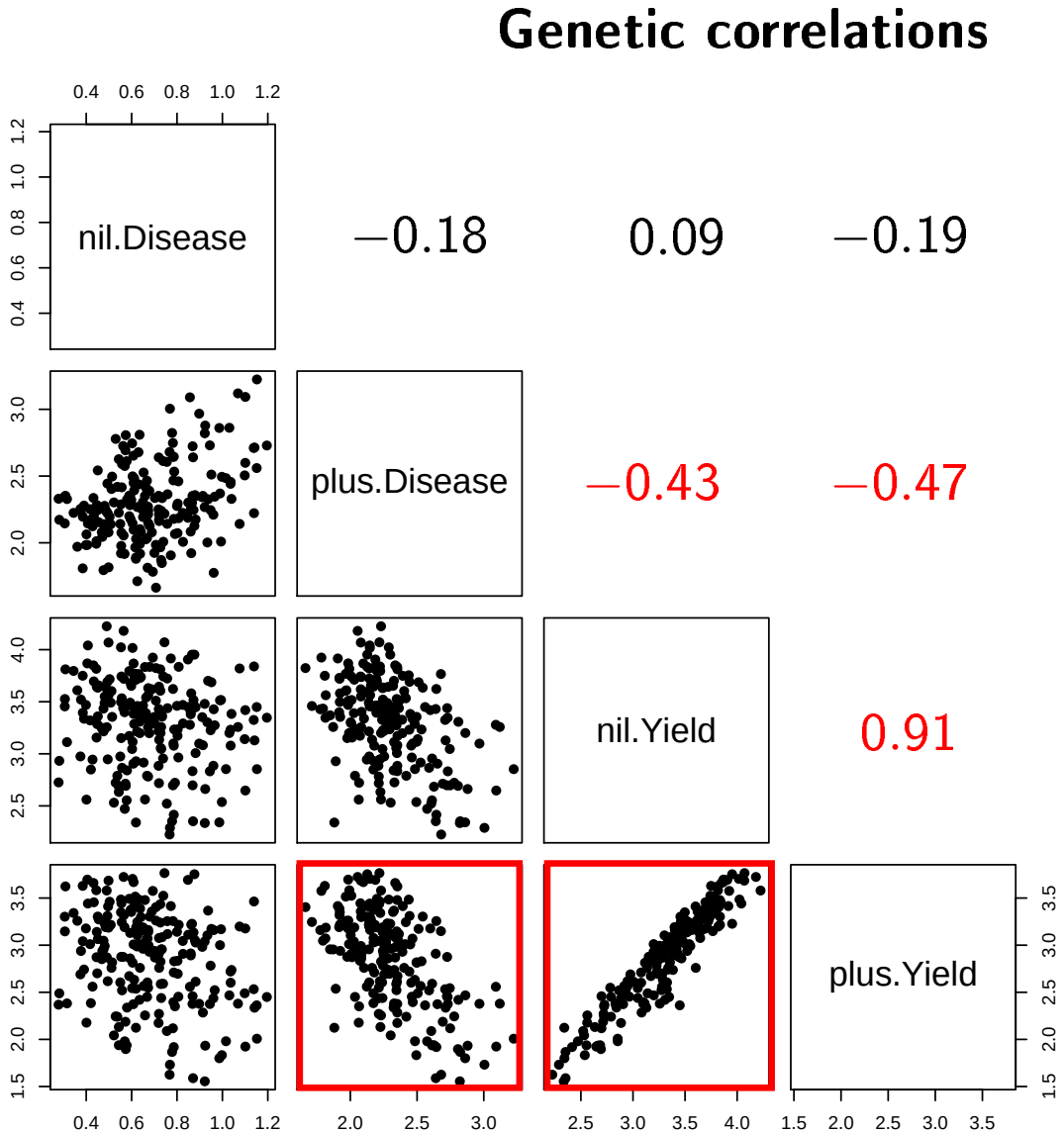
Trait by Inoculum means and genetic variances estimated from the model.

Trait	Inoculum	Mean	Genetic variance
Disease	nil	0.68	0.084
(sqrt(cm))	plus	2.29	0.128
Yield	nil	3.32	0.197
(kg/ha)	plus	2.88	0.270

Results
Genotype predictions (BLUPs)



Results
Genotype predictions (BLUPs)



The bivariate linear mixed model

We focus on the genotype blups, and use the regression structure inherent in the bivariate model to interpret the results.

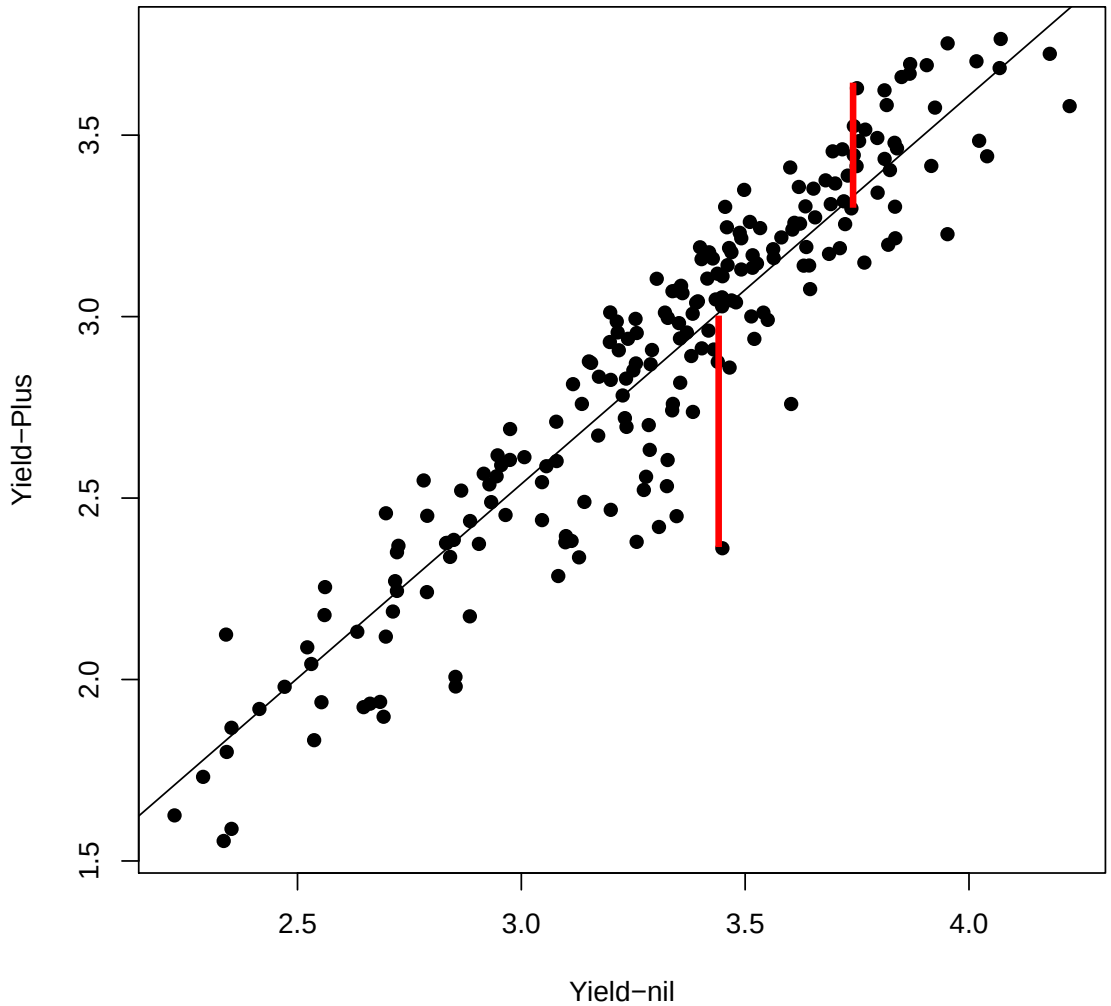
For Yield, plus vs nil blups

$$\mathbf{u}_{g_{Y_p}} = \beta_Y \mathbf{u}_{g_{Y_n}} + \boldsymbol{\epsilon}_Y, \qquad \beta_Y = \frac{\sigma_{g_{Y_{np}}}}{\sigma^2_{g_{Y_n}}}.$$

Define yield responsiveness as $\boldsymbol{\epsilon}_Y$, where responsiveness has zero covariance with yield potential, $\mathbf{u}_{g_{Y_n}}$.

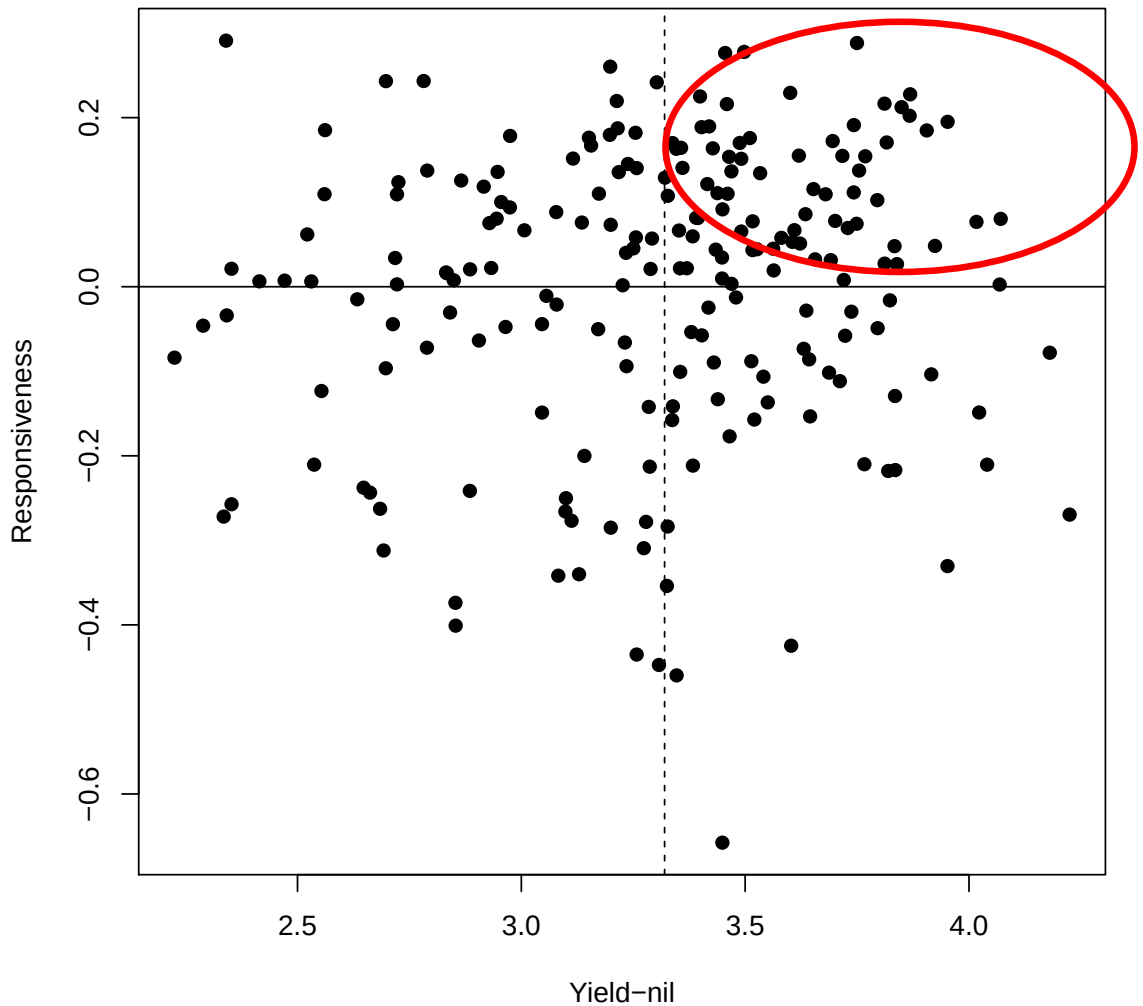
Results
Genotype predictions (BLUPs) from the Crown Rot Tolerance trial

Yield responsiveness, ϵ_Y



Yield responsiveness of genotypes

Genotypes with high yield potential and responsiveness under disease pressure



The bivariate linear mixed model

Again, we use the regression structure inherent in the bivariate model to interpret the results.

For Plus disease plots, yield vs disease blups

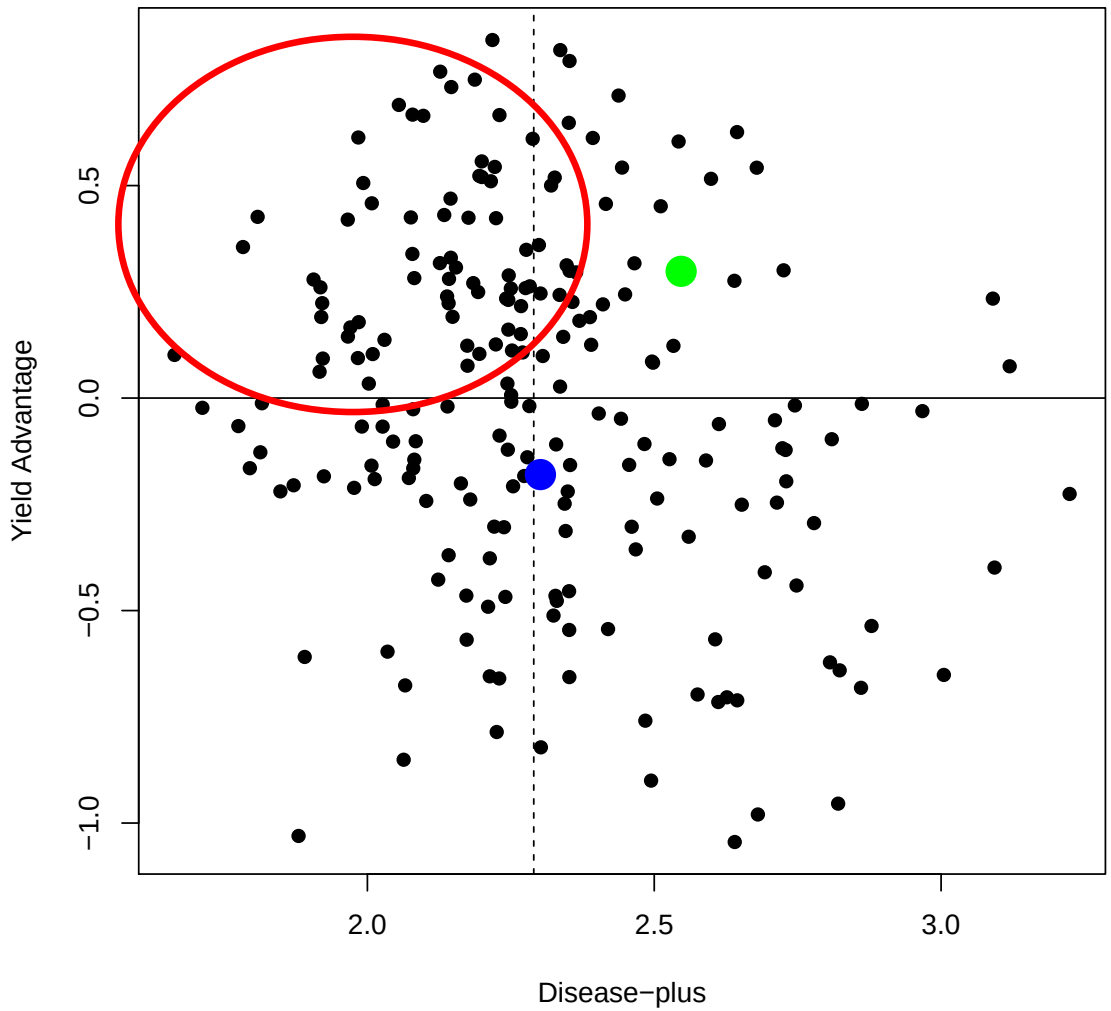
$$\mathbf{u}_{g_{Y_p}} = \beta_P \mathbf{u}_{g_{D_p}} + \boldsymbol{\epsilon}_P, \qquad \beta_P = \frac{\sigma_{g_{Y_p D_p}}}{\sigma^2_{g_{D_p}}}.$$

Define yield advantage under disease as $\boldsymbol{\epsilon}_P$, where yield advantage has zero covariance with disease severity under inoculation, $\mathbf{u}_{g_{D_p}}$.

Yield advantage of genotypes under disease

Combined selection for tolerance and resistance

Yield advantage under disease, ϵ_D



A bivariate statistical analysis of yield and disease

Extensions to this analysis

- We have extended this model to an analysis across multiple trials (environments).
- We have included a genetic covariance relationship through the pedigree of the genotypes.
- This will be extended to a genomic relationship matrix as marker data is available next year.

Acknowledgements

The National Crown Rot Initiative team

- Phillip Davies, Cassy Percy, Chunji Liu
- Field staff at Plant Breeding Institute Narrabri
- Grains Research and Developoment Corporation