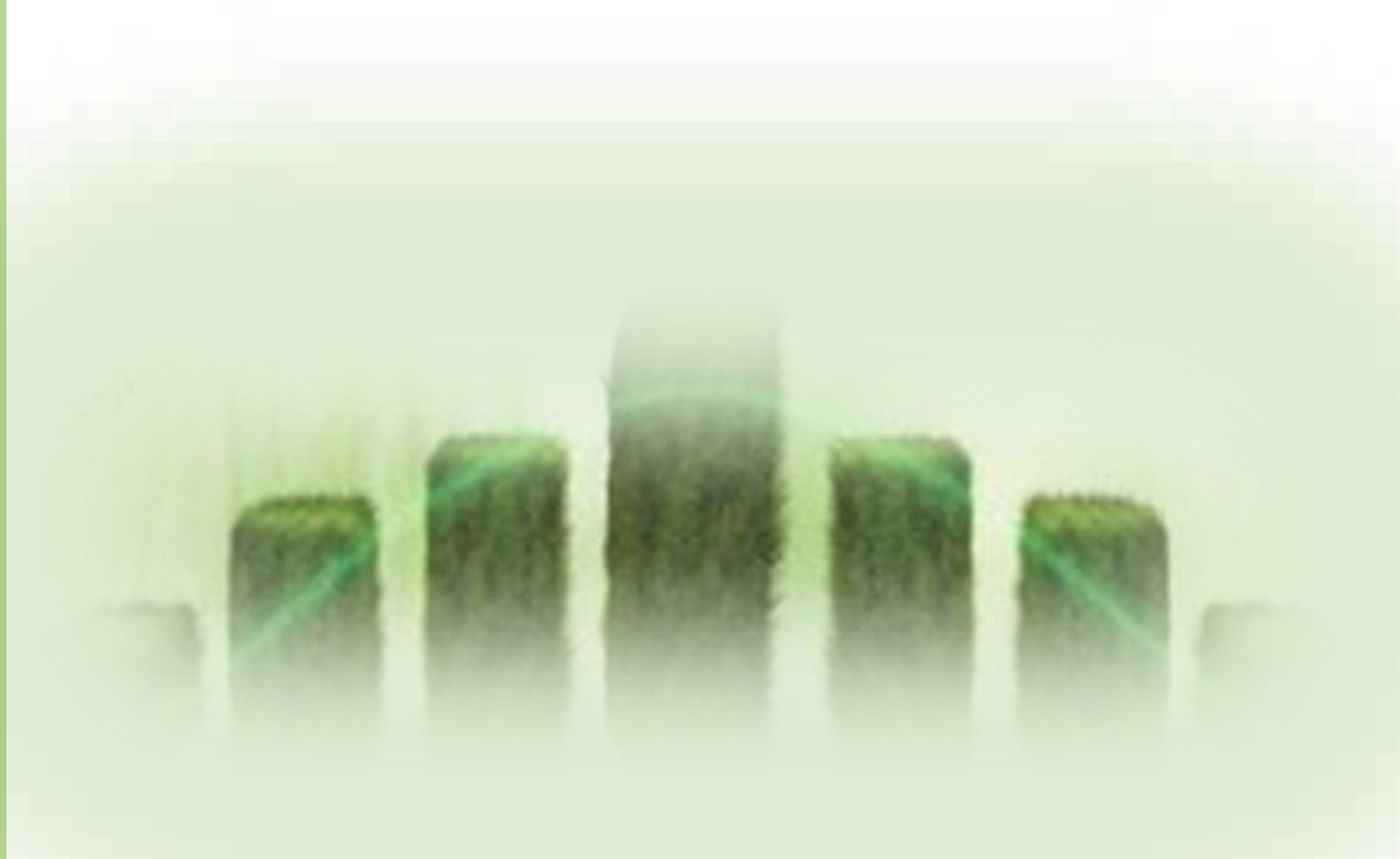
Conditional Inference Trees and ANOVA comparison

Iain Proctor, BASF SE, 29th May 2017



„Same procedure as every year“

- ANOVA is a standard procedure, but unfortunately not always straightforward for everyone
- Practice for exploratory models follows protocol for ANOVA for registration assessments:
  - Choice of transformation
  - Normally distributed residuals
  - Understanding of significance level
  - Choice of post hoc test
  - Interpretation of the result

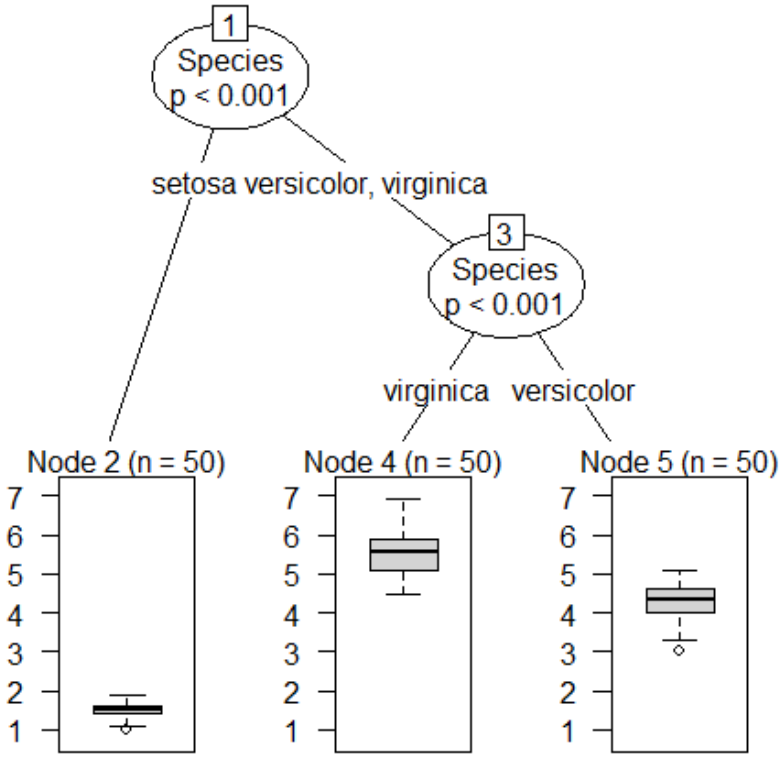
0.48 abc
0.19 abc
0.30 abc
0.68 abc
0.83 ab
0.68 abc
0.76 abc
1.00 a
0.34 abc
0.04 bc
0.09 bc
0.24 abc
0.30 abc
0.00 c

Needs

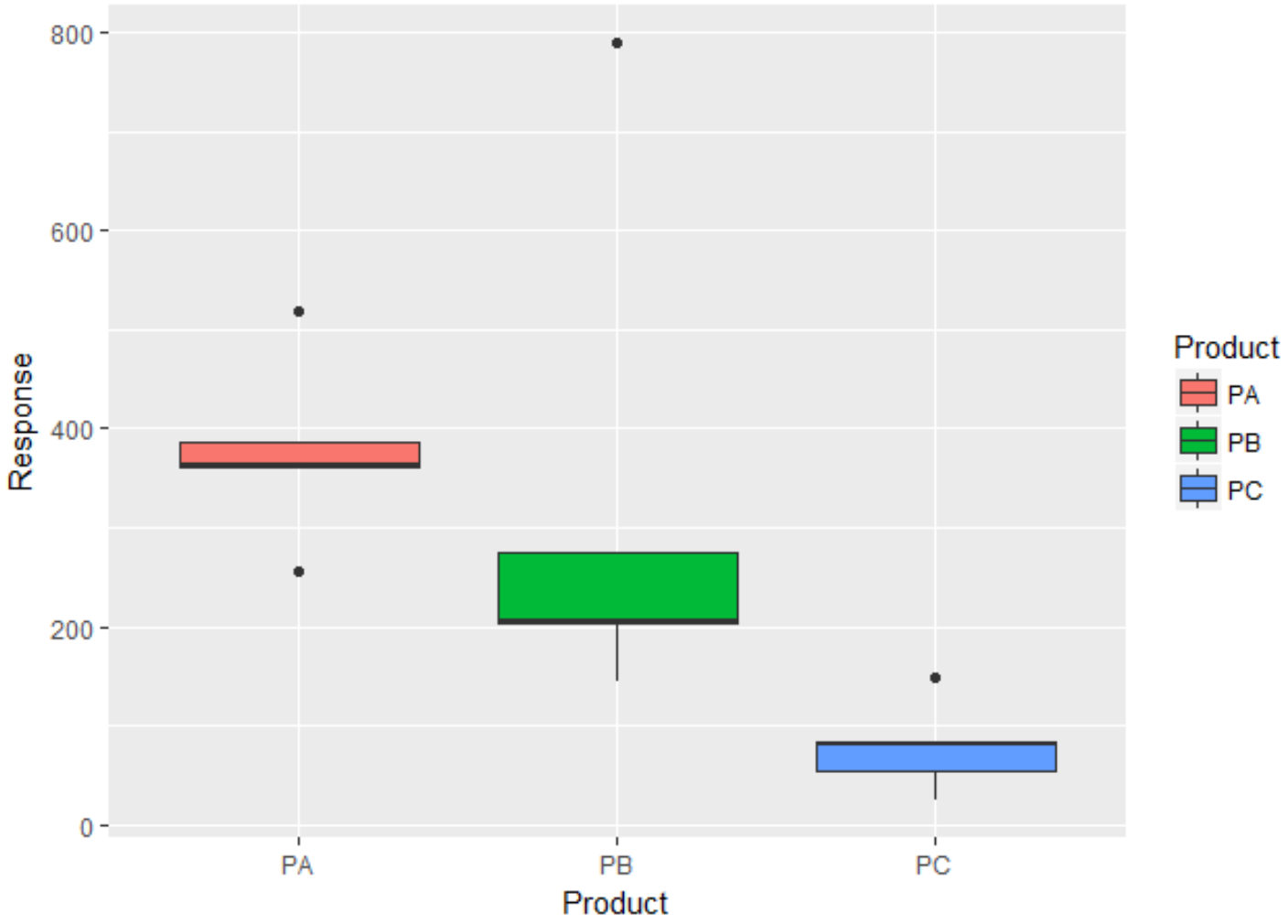
- Problem
  - Choice of transformation
  - Normally distributed residuals
  - Understanding of significance level
  - Choice of post hoc test
  - Interpretation of the result
- Requirement
  - No transformation necessary
  - No residual check required
  - Obvious answer to significance of factor
  - Defined test
  - Easily understandable result

Conditional inference trees

- *partykit* package in R
- Iris data: are species different in terms of petal length?
  - Petal length ~ Species
- Local optimum chosen for split in data
- Where there is more than one covariate, the most significant covariate in model ist chosen
- Minimum adequate model returned

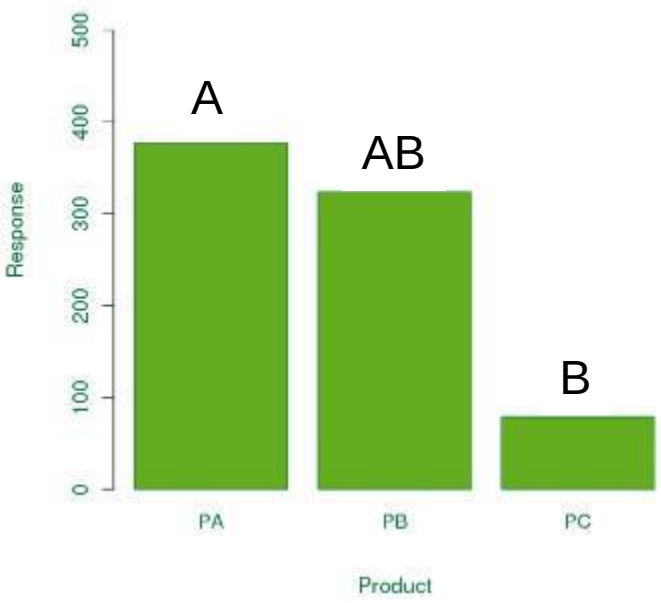


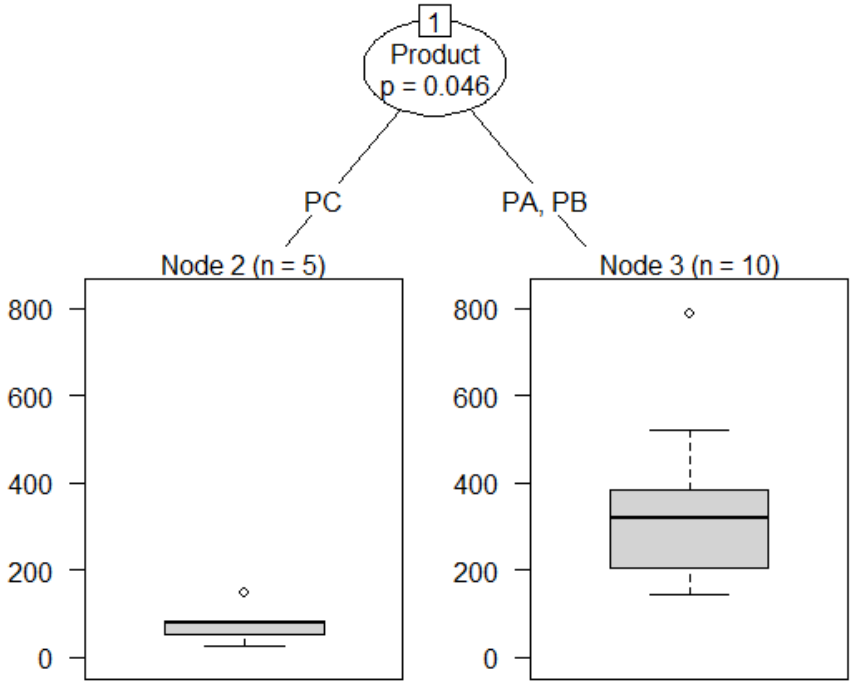
Example data



Comparison of output

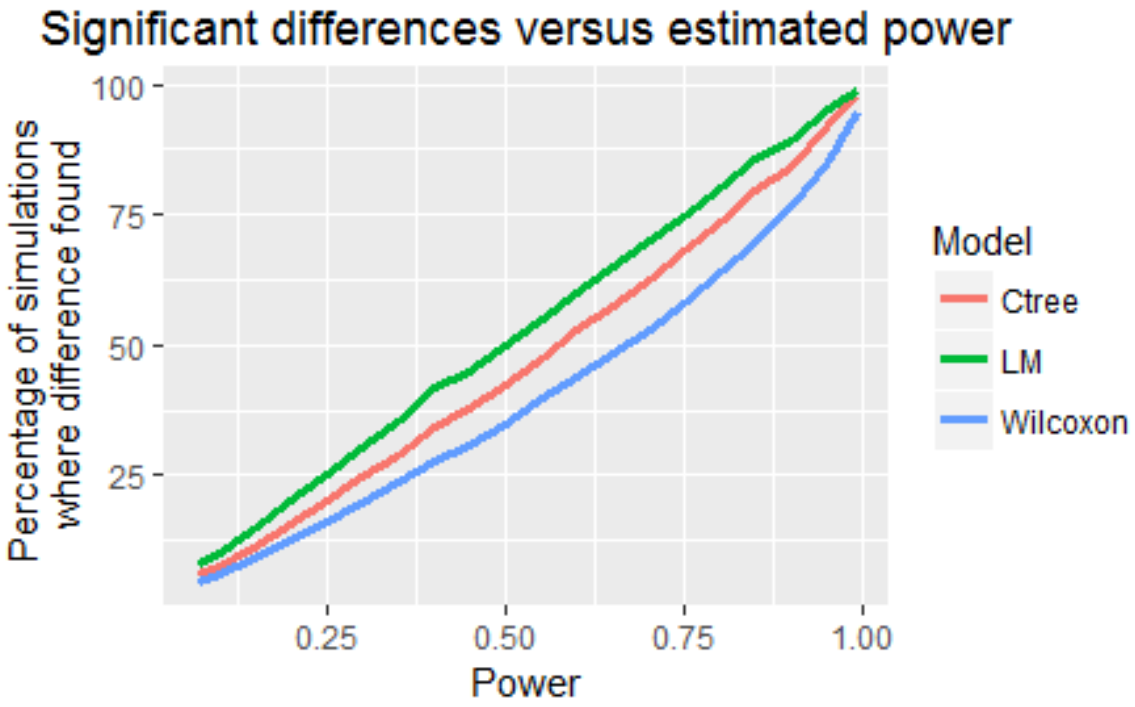
Factor	Df	SS	MS	F-value	P-value
Product	2	253111	126556	47.046	0.03101
Residuals	12	322807	26901		





Simulation comparison

- Data simulated for two treatment groups with varying power
- Linear model, Ctree and Wilcoxon Rank Sum Test conducted on data



Real data

- Measurements from the BASF field trial database tested using ANOVA and Ctree, using response as dependent variable and treatment as independent covariate
- 8000 models tested:

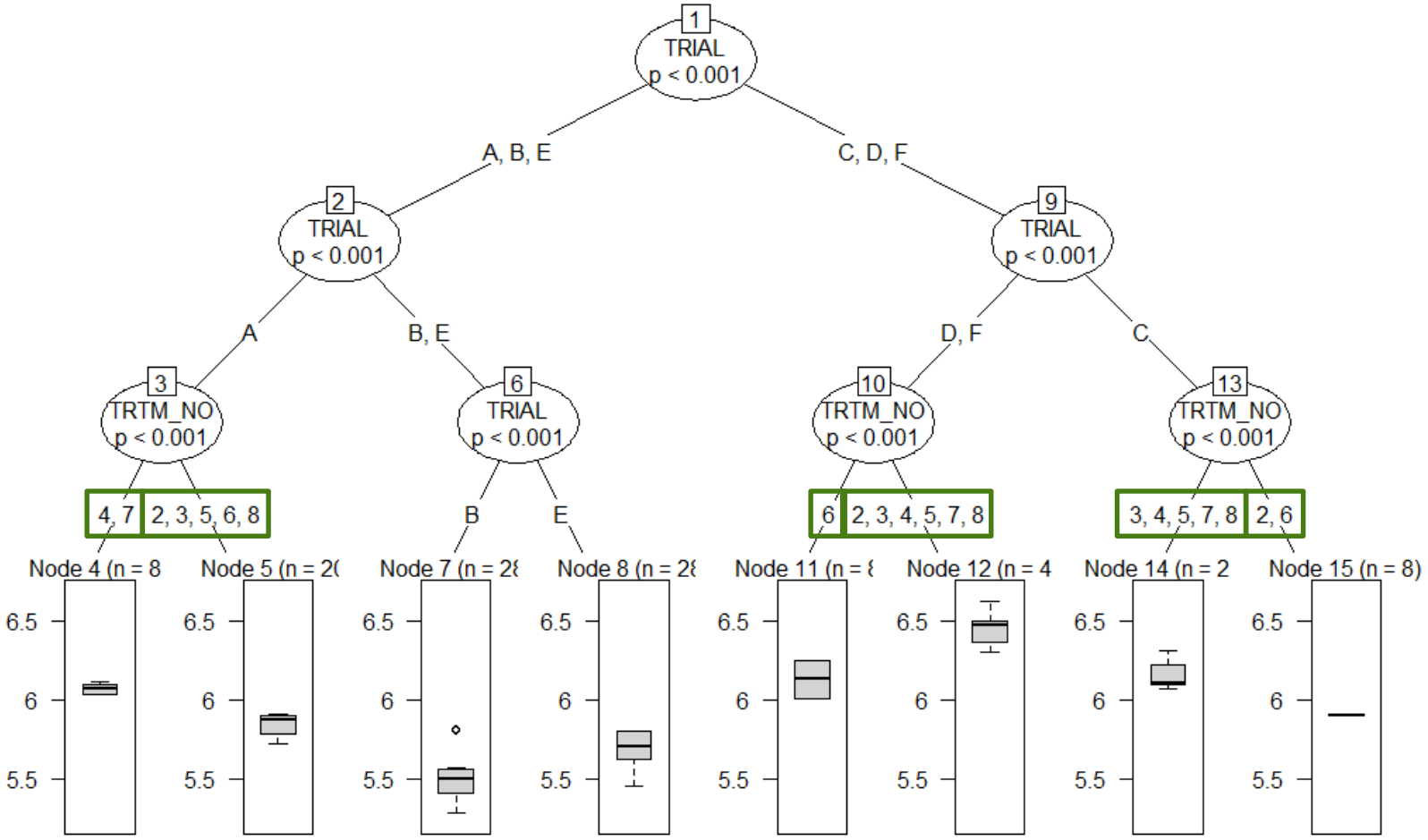
Model	Sig. Difference	No sig. difference
ANOVA	2456	5544
Ctree	2376	5624

- 82 models where ANOVA returned a significant difference and Ctree did not
- 2 models where Ctree returned a significant difference and ANOVA did not

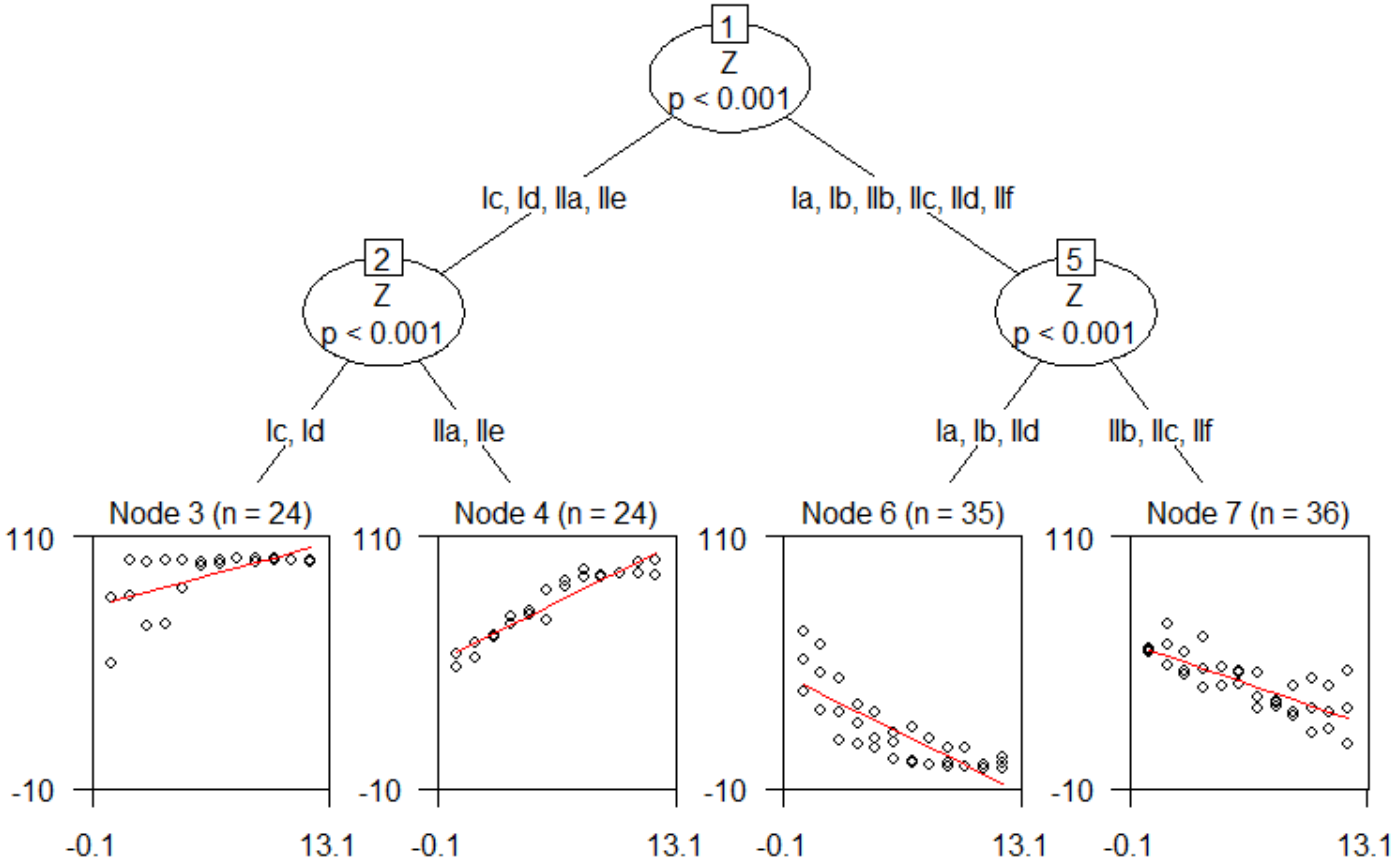
Potential drawbacks

- Complexity of post hoc test result can be lost in conditional inference tree result
- Change of result based on inclusion of third parties
- In agricultural field trials, such a result can be justified (or avoided) by only including those treatments in the model which were part of the original trial design.
- Reappearance of covariate in tree (Covariate A, followed by B followed by A again) could be considered as confusing as an interaction effect to explain

Extension to 2 factors – trial series



Extension to model fit at each node



Conclusion

- Conditional inference trees are useful for exploratory analysis, prior to registration
- Requires little statistical knowledge for interpretation of the resultant plot
- Presence of more than one end node is closely correlated with siginificance in an ANOVA with treatment; improved power relative to Wilcoxon Rank Sum
- Minimum adequate model results in simplified effects and interactions
- Can be combined for more complex assessment with regression model at the end of each node

We create chemistry