Evaluation of threshold method for early distinctness decisions

Adrian Roberts and Ian Nevison
Biomathematics and Statistics Scotland

Eleventh Working Seminar on Statistical Methods in Variety Testing

Outline

- Background
- Objective
- Formulation
- Heteroscedasticity
- Proposed solution
- Evaluation on real data

Funded by the Scottish Government
Thanks to Tom Christie of Science and Advice for Scottish Agriculture

Introduction

Plant Breeders' Rights (PBR) – protection for plant varieties

International Convention for the Protection of New Varieties of Plants

- 75 members
- International Union for the Protection of New Varieties of Plants (UPOV)
- To be granted rights, a variety must be distinct, uniform and stable (DUS)

DUS

Distinct: differs from all other known varieties by one or more important botanical characteristics, such as height, maturity, colour, etc.

Uniform: plant characteristics are consistent from plant to plant within the variety

Stable: plant characteristics are genetically fixed

DUS assessment

New varieties compared to existing varieties in trials over two or three cycles (= *years*)

- Usually one location
- Existing varieties thought to be reasonably similar to candidate are planted

Characteristics defined in advance for a crop

- Generally not performance characters
- Qualitative or quantitative
- Quite numerous (e.g. Pea ~ 20)

DUS assessment

Evaluation is characteristic-by-characteristic

- not multivariate

Must be distinct from every variety but only for one characteristic in each

Needs to be uniform in all characteristics considered

Distinctness for quantitative characteristics

COYD – Combined Over-Years Distinctness criterion

- Statistically-based criterion for measured, quantitative characters
- Based on ANOVA on trial x variety means
- Univariate - applied character-by-character
- Candidate distinct if significantly different from all other varieties for at least one characteristic at the prescribed significance level.

Objective of this work

Based on results from the first year, identify varieties that are highly likely to be distinct from a candidate after two years (using COYD)

No need to compare these in second year

Potential to reduce trial size

Decision making after one year

We wish to:

Set a threshold for declaring "distinct" in first year

- Based on prediction of the 2-year COYD verdict
- Use data from the first year of tests plus historical data

Analytical solution preferred (easier implementation)

The COYD criterion

Mean measurements, x_{ij} , are made on the characteristic of interest for each variety, i, and year, j.

We are interested in varieties A and B

Let $$d_j = x_{Aj} - x_{Bj}$$ *Difference in year j*

and $$D = \frac{d_1 + d_2}{2}$$ *Difference over both years*

The COYD criterion

COYD criterion says *A* and *B* distinct if:

$$\left|\frac{D}{s_{12}}\right| \geq t_{1-p/2,\nu_{12}}$$

s_{12} is sq root estimated residual error from COYD ANOVA

p is COYD significance level

v_{12} is degrees of freedom from COYD ANOVA.

The COYD criterion

COYD criterion says *A* and *B* distinct if:

$$\left|\frac{D}{s_{12}}\right| \geq t_{1-p/2,\nu_{12}}$$

s_{12} is estimated residual std dev from COYD ANOVA

p is COYD significance level

v_{12} is degrees of freedom from COYD ANOVA.

If we want to predict this decision based on only one year, we need to predict $\mathbf{d_2}$ *&* $\mathbf{s_{12}}$*. We can do this based on* d_1 *and an analysis of the historical data.*

Assumptions

Assume errors for x_{ij} are iid (normal):

- So straightforward to specify the predictive distribution for D/s_{12}
- then can work out thresholds etc.
- But these assumptions are strong

Assumptions

Previously reviewed the importance of this set of assumptions:

- How relevant is it to real data (in a pea example)
- Effect of departures:
  - Skewness
  - Kurtosis
  - Heterogeneity of residual variance by variety or year

⇨ Main concern is heterogeneity by year

  - Common & important

Proposed way to deal with heteroscedasticity

Allow the variances of the observations vary from year-to-year.

So let

$$x_{ij} \sim N(\mu_{ij}, \sigma_j^2)$$

such that the precisions come from a gamma distribution

$$\lambda_j = \frac{1}{\sigma_j^2} \sim Ga(\alpha, \beta)$$

Estimating Gamma distribution parameters

Estimate α and β from historical data using an approximation (Minka, 2002)

Let $$\delta = \ln\left(\frac{1}{n}\sum \frac{1}{s_j^2}\right) - \frac{1}{n}\sum \ln\left(\frac{1}{s_j^2}\right),$$

where s_j^2 is the residual variance for year j, and n is the number of years.

Then $$\hat{\alpha} \approx \frac{3 - \delta + \sqrt{(\delta - 3)^2 + 24\delta}}{12\delta}$$

and $$\hat{\beta} = \frac{\hat{\alpha} n}{\sum \frac{1}{s_j^2}}$$

Calculating approximate thresholds

We have the distributions for both *D* and s_{12}

Using Taylor series, we can derive approximations for the expectation and variance of the ratio

- *assumes independence – but this seems to be okay*

$$E\left[\frac{D}{s_{12}}\right] \approx \left(1 + \frac{3}{8}\frac{v_{12} + 2\hat{\alpha} - 2}{v_{12}(\hat{\alpha} - 2)}\right) d_1 \sqrt{\frac{\hat{\alpha} - 1}{\hat{\alpha}} \frac{\sum \frac{1}{s_j^2}}{n}}$$

$$V\left[\frac{D}{s_{12}}\right] \approx \frac{1}{2}\left(1 + \frac{d_1^2}{2}\frac{(\hat{\alpha} - 1)}{\hat{\alpha}}\frac{\sum \frac{1}{s_j^2}}{n}\frac{(v_{12} + 2\hat{\alpha} - 2)}{v_{12}(\hat{\alpha} - 2)}\right)$$

Calculating approximate thresholds

We have the distributions for both *D* and s_{12}

Using Taylor series, we can derive approximations for the expectation and variance of the ratio

- *assumes independence – but this seems to be okay*

$$E\left[\frac{D}{s_{12}}\right] \approx \left(1+\frac{3}{8}\frac{v_{12}+2\hat{\alpha}-2}{v_{12}(\hat{\alpha}-2)}\right) d_1 \sqrt{\frac{\hat{\alpha}-1}{\hat{\alpha}}\frac{\sum\frac{1}{s_j^2}}{n}}$$

$$V\left[\frac{D}{s_{12}}\right] \approx \frac{1}{2}\left(1+\frac{d_1^2}{2}\frac{(\hat{\alpha}-1)}{\hat{\alpha}}\frac{\sum\frac{1}{s_j^2}}{n}\frac{(v_{12}+2\hat{\alpha}-2)}{v_{12}(\hat{\alpha}-2)}\right)$$

Propose to fit a normal distribution to ratio with above mean and variance (works better than Student t-distribution)

Calculating approximate thresholds

Define threshold as the value of d_1 such that

$$p_D = \Pr\left(\left| \frac{d_1 + d_2}{2s_{12}} \right| \geq t_{1-p/2,\nu_{12}} \;\middle|\; d_1, \underline{X} \right)$$

where p_D is the probability of passing the COYD criterion

Simulations to see how well these thresholds work given the approximations

Evaluation of approximations by simulation

Presented at 10th Working Seminar

+ *Roberts, Nevison, Christie (2016) Journal of Agricultural Science, 154: 1317-1326*

Conclusion: approximation is reasonable within likely usage

How useful is this method in practice?

Test with real data:

Data from Finland and the United Kingdom

Calculate thresholds based on this historical data

Probability p_D: 0.95, 0.98, 0.99

Compare decisions based on thresholds after one year with 2 year COYD decisions

Data sets

Country	Crop	Number of cycles	Probability level for COYD	Number of characters used here	Overall number of varieties	Overall number of candidates
Finland	Meadow fescue	12	0.01	5	64	23
Finland	Red Clover	11	0.01	6	39	10
Finland	Timothy	11	0.01	6	100	9
United Kingdom	Perennial ryegrass	11	0.01	16	232	146
United Kingdom	Pea – semi leafless	19	0.02	10	887	275
United Kingdom	Pea – conventional	20	0.02	12	405	58

UK pea Thresholds
Semi-Leafless Group

UPOV no	Characteristic	Mean COYD criterion	Threshold with p_D=0.95	Threshold with p_D=0.98	Threshold with p_D=0.99
5	Stem: number of nodes up to and including first fertile node	0.86	1.81	2.73	4.13
15	Stipule: length (mm)	10.58	17.90	20.91	23.38
16	Stipule: width (mm)	6.72	11.15	12.84	14.18
22	Petiole: length from axil to first leaflet or tendril (mm)	12.26	21.31	25.16	28.38
28	Flower: width of standard (mm)	2.30	4.18	5.13	5.99
34	Peduncle: length from stem to first pod (mm)	19.49	33.46	40.00	45.63
37	Pod: length (mm)	5.91	9.79	11.33	12.56
38	Pod: width (mm)	0.96	1.59	1.82	2.00
46	Pod: number of ovules	0.45	0.77	0.91	1.03

How useful is this method in practice?

Next step: assessing performance

- Do we get first cycle decisions correct?
- What reductions could be achieved?

How useful is this method in practice?

Apply calculated thresholds to the data sets

compare first cycle decisions using thresholds with 2-cycle COYD decisions

False positive rate for each characteristic:

first-cycle threshold distinct: COYD non-distinct

False negative rate for each characteristic:

first-cycle threshold non-distinct: COYD distinct

How useful is this method in practice?

Apply calculated thresholds to the data sets

compare first cycle decisions using thresholds with 2-cycle COYD decisions

False positive rate for each characteristic:

first-cycle threshold distinct: COYD non-distinct

False negative rate for each characteristic:

first-cycle threshold non-distinct: COYD distinct

Want very low false positive rate to avoid poor decisions
but need low false negative rate to make it worthwhile

How useful is this method in practice?

NOTES OF CAUTION:

Real data: reference varieties may have been removed after first cycle

false negative rate over-estimated?

Decisions are made over the set of characteristics

Here we only included characteristics with thresholds

May be other characteristics (qualitative) that can contribute to decisions

UK pea Thresholds
Semi-Leafless Group

Characteristic No.	False positives (%)			False negatives (%)		
	p_D=0.99	p_D=0.98	p_D=0.95	p_D=0.99	p_D=0.98	p_D=0.95
5	0.0	0.0	0.4	85.8	64.0	40.0
15	0.3	0.7	1.8	86.0	78.4	65.2
16	0.5	0.8	2.1	74.2	66.3	54.1
22	0.1	0.4	1.4	89.0	81.8	69.1
28	0.0	0.3	1.0	89.0	81.3	66.0
34	0.0	0.1	0.8	85.1	76.8	61.6
37	0.0	0.2	0.7	79.5	73.3	61.7
5	0.2	0.6	1.6	76.5	67.7	56.0
46	0.1	0.4	1.4	63.8	55.3	41.7
57	0.0	0.1	0.6	61.1	50.1	37.3

UK pea Thresholds

Semi-Leafless Group

Characteristic No.	False positives (%)			False negatives (%)		
	p_D=0.99	p_D=0.98	p_D=0.95	p_D=0.99	p_D=0.98	p_D=0.95
5	0.0	0.0	0.4	85.8	64.0	40.0
15	0.3	0.7	1.8	86.0	78.4	65.2
16	0.5	0.8	2.1	74.2	66.3	54.1
22	0.1	0.4	1.4	89.0	81.8	69.1
28	0.0	0.3	1.0	89.0	81.3	66.0
34	0.0	0.1	0.8	85.1	76.8	61.6
37	0.0	0.2	0.7	79.5	73.3	61.7
38	0.2	0.6	1.6	76.5	67.7	56.0
46	0.1	0.4	1.4	63.8	55.3	41.7
57	0.0	0.1	0.6	61.1	50.1	37.3

Over Characteristics

Data set	False positives (%)			False negatives (%)		
	p_D=0.99	p_D=0.98	p_D=0.95	p_D=0.99	p_D=0.98	p_D=0.95
Meadow fescue	0.0	0.7	2.7	95.2	87.3	66.4
Red Clover	0.0	0.0	4.8	100.0	73.5	37.1
Timothy	0.1	0.1	1.0	96.2	90.1	72.0
Perennial ryegrass	0.2	1.0	7.7	69.2	48.3	22.6
Pea – semi-leafless without groups	0.5	0.5	8.1	45.6	29.7	15.0
Pea – conventional	0.0	0.0	2.4	85.2	71.4	26.3

Over Characteristics

Data set	False positives (%)			False negatives (%)		
	p_D=0.99	p_D=0.98	p_D=0.95	p_D=0.99	p_D=0.98	p_D=0.95
Meadow fescue	0.0	0.7	2.7	95.2	87.3	66.4
Red Clover	0.0	0.0	4.8	100.0	73.5	37.1
Timothy	0.1	0.1	1.0	96.2	90.1	72.0
Perennial ryegrass	0.2	1.0	7.7	69.2	48.3	22.6
Pea – semi-leafless without groups	0.5	0.5	8.1	45.6	29.7	15.0
Pea – semi-leafless with groups	0.8	0.8	9.4	65.7	45.9	24.2
Pea – conventional	0.0	0.0	2.4	85.2	71.4	26.3

Sensitivity

Sensitivity to data set:

- For conventional pea group, looked at effect of restricting data set to varieties with 2,3, 4, 5 or 6 cycles present
- No relationship with number of varieties present
- But threshold at 99% much more sensitive

Sensitivity

Characteristic No.	CV		
	p_D=0.99	p_D=0.98	p_D=0.95
9	0.36	0.14	0.03
10	0.30	0.13	0.06
11	0.15	0.02	0.04
14	0.30	0.15	0.18
15	0.36	0.09	0.03
16	0.17	0.06	0.09
28	0.23	0.15	0.18
34	0.28	0.10	0.02
37	0.08	0.04	0.01
38	0.25	0.10	0.09
46	0.25	0.06	0.06
57	0.16	0.06	0.02

Summary of results

Quality of thresholds depends on:

- Size of historic data set
- Number of cycles
- Number of reference varieties
- Number of varieties in common between cycles

Utility of method depends on size of current trials

- Smaller trials lead to larger thresholds (esp 99%)

Sensitivity to data set:

- Threshold at 99% much more sensitive
- Need to examine ways to make this more robust

Conclusions

Method is most applicable to crops with large reference collections and where current trial sizes are large

Utility will depend on crop and DUS trialling system

- Works for pea in UK – measured characteristics in combination with groups
- May also work where similar varieties are planted together in second cycle or in combination with markers
- Combination with GAIA (Geves)?
- Recently received oilseed rape data (from UK and Denmark) – will evaluate this summer

Code developed in R