Protocol for determining infestations of whitefly

William Godsoe, Emiliano Veronesi

and Stephen Goldson

Faculty of Agriculture and Life Sciences,

Lincoln University

July 2022

Report and protocol for “A Lighter Touch”

Contents

Suggested protocol 3

Section 1: Estimation of precision of sampling of whitefly populations in greenhouse tomatoes 4

Accuracy of estimation weekly scores based on sampling size. Blue line shows weekly over- and under-estimation of the mean score based on sample size .....

Decision-making figures as white fly population build up, typically week-by-week. .....

Section 2: Analysis of our ability to detect incursions of rare insects such as Tomato potato psyllid (TPP) 7

Appendix 1: additional background analyses for section 1 9

Part 1: Formula for how many samples will be needed to compute average scores 9

Details of calculation 10

Evaluation 10

Part 2: an outbreak where flies are distributed at random 12

Part 3: simulations of different thresholds when whiteflies are random 13

Part 4: simulations when whitefly outbreaks are spatially clumped 14

Part 5: simulations of different thresholds when flies are clumped 15

Part 6: Testing if early whitefly outbreaks are spatially clumped 16

Part 7: Examples of clumping when scores are large 17

Part 8: Scores are close to random when the number of flies is small 18

Part 9: Analysis including data from reporoa 20

Suggested protocol

We suggest the following protocol for identifying the overall strength of infestations of pests such as whitefly:

- Score number of whiteflies per plant (see below for an example scale)
- Do 50 samples that are representative of the glasshouse.
    - This can be individual plants
    - Find a convenient way to collect samples.
    - Make sure your sampling is representative of the entire glasshouse.
    - Since outbreaks can grow rapidly, it is good to sample often, for example once per week.
- The first few samples are most important. There will be diminishing returns for additional samples.
- Charts in section 1 show how much variability to expect for the average score of your sample. This variability depends on the average score.

Level	White fly numbers
0	0
1	<5
2	6-10
6	11-15
8	16-20
10	>20

Section 1: Estimation of precision of sampling of whitefly populations in greenhouse tomatoes

The scope of this work is to evaluate sampling protocols for detecting whiteflies in tomato growing facilities.

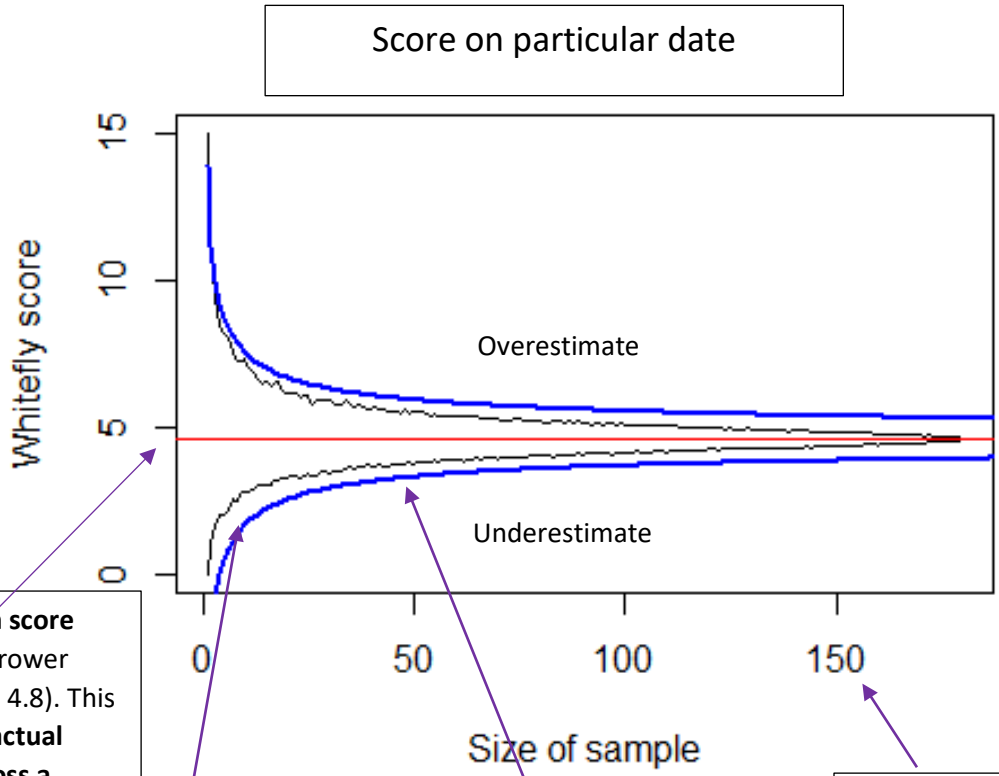
Commonly collected data is a numerical score, based on impression, of the extent of whitefly infestation on individual plants. These scores are typically from between 0 and 6 with 6 representing high infestations. Based the distribution of scores in field data (combined with statistical theory, see attached appendix), it is possible to determine optimum whitefly sampling intensities to achieve reliable information to for pest-control decision-making.

The model presented in this work provides a tool to prevent sub-optimal sampling. There are basically two types of sub-optimal sampling. The first is not to sample sufficiently such that the population estimates are very broad and therefore too approximate to be useful. At the other end of the spectrum, additional sampling really does not add usefully to the precision of the estimates. This is sometimes called the "law of diminishing returns". For example, sometimes sampling more than 50 times can lead to little increased precision depending on the pest population. These considerations must also factor in the prevailing whitefly population in a greenhouse at any given time.

Rather than deal with quite a few equations and calculations it is better to use decision-making diagrams; the use of these is described below along with reference figures for use. As the mean score in a facility goes up, samples become increasingly variable. To reflect this, on page 3 of this report we present decision diagrams for mean scores between 1 and 6.

White fly Infestation scores beyond 6 will probably suggest that, given the population present, sampling variation will matter less as the decision to do something will be very obvious.

Accuracy of estimation weekly scores based on sampling size. Blue line shows weekly over- and under-estimation of the mean score based on sample size



This is the **mean score** resulting from grower sampling (about 4.8). This represents the **actual** mean **score across a facility**

Here is a sample of about 10. The mean score here is very approximate and far from the real number of whiteflies in the house (red line). This is an example of **under-sampling**.

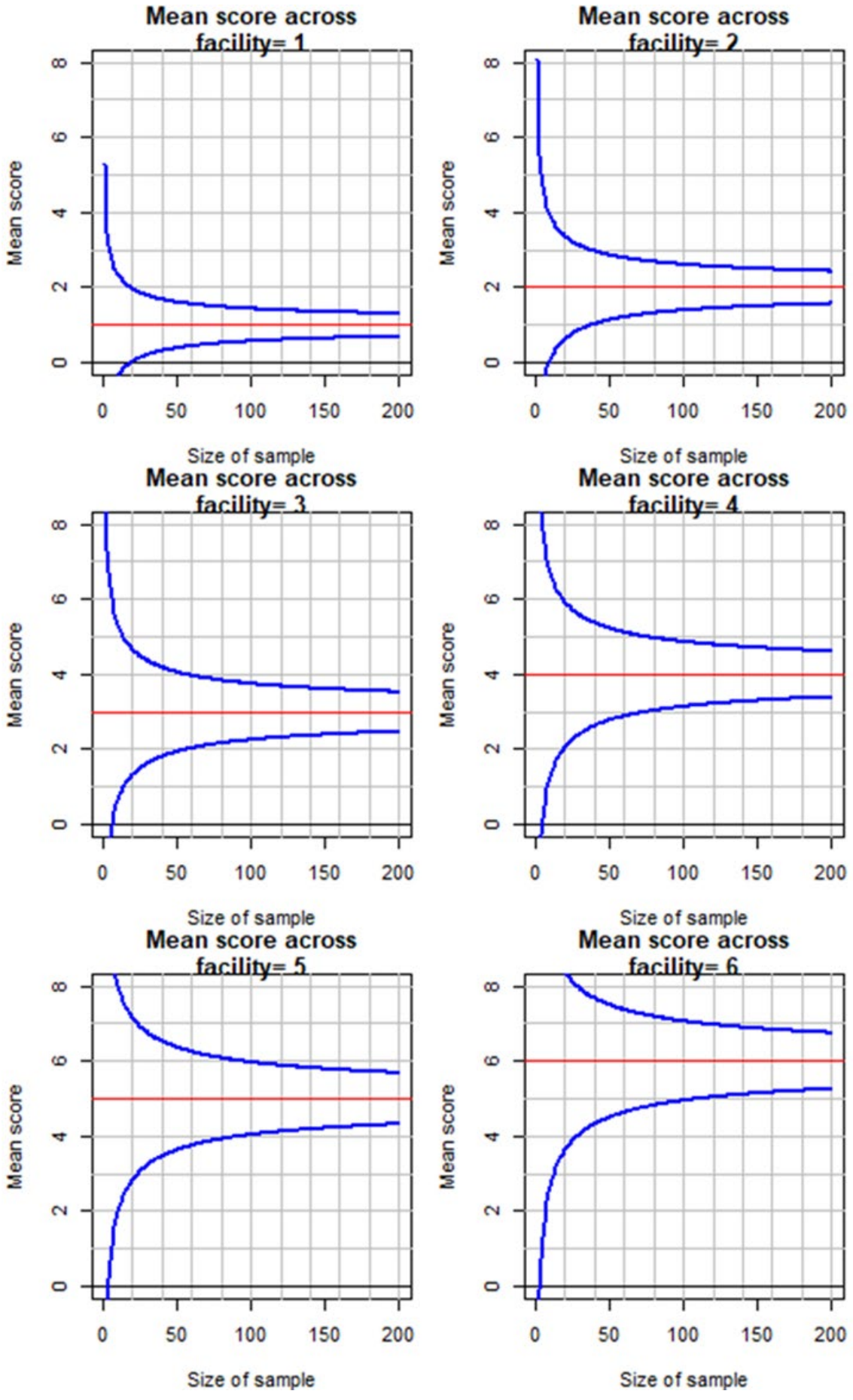
Here is a sample of about 50. Good estimate as the blue line is fairly close to the red one.

Here is a sample of about 150. Not much better than 50 as per blue line (diminishing returns). This is an example of **oversampling**.

Notes:
The blue line is the mathematical model the black squiggly line is as actually measured scores.
Grid lines left off this diagram for simplicity.

Decision-making figures as white fly population build up, typically week-by-week.

Point of inflection of blue line suggest the point of max sampling efficiency



Section 2: Analysis of our ability to detect incursions of rare insects such as Tomato potato psyllid (TPP)

The goal of this section is try to evaluate the chance of detecting an incursion of an insect such as TPP when it is very rare. To do this I'm going to break the problem down into the events that are needed for an incursion to be detected, then combine information on each of these events to understand the chance that an outbreak will be detected.

The previous section work highlighted how increasing sampling improves your ability to detect the average intensity of an insect outbreak. That previous report is most useful when an outbreak is already established, and we just want to know the average intensity. This work highlights what is needed to detect insects when they are extremely rare.

There are three major ways that TPP can be missed. At any given time TPP will be absent from many glasshouses. Because of this, some surveys will fail to detect TPP. Within a survey, some portions of the glasshouse will not be sampled. As a result, some surveys will fail to detect TPP. When the correct part of the glasshouse is surveyed TPP may be missed, and as a result, some surveys will fail to detect TPP. We can get a sense of the chance of detecting TPP by combining information on each of these three processes.

The probability that, in a given survey, you will detect an incursion of a rare insect is:

$$Chance\ to\ detect\ incursion = Chance\ incursion\ present \times chance\ surveyed \times chance\ noticed$$

The chance that a survey will detect an incursion can be small because it requires three events each of which could unlikely. For example, lets say that an incursion is present at 1 in 10 glass houses, a survey examines one tenth of the glasshouse at a time and the incursion is noticed ¼ of the time that it is found, this leads to:

$$Chance\ to\ detect\ incursion = \frac{1}{10} \times \frac{1}{10} \times \frac{1}{4} = 0.0025$$

In other words, you will detect 2.5 incursions for every 1000 surveys of the glasshouse.

We have attached the document "tool for survey detection" as a way to calculate examples, using information about a particular facility. To do this, enter the chance an incursion is present, the chance it is surveyed and the chance it is noticed. Each of these chances will be a number between 0 (representing something that is impossible) and 1 (representing something that is guaranteed).

Present	Survey	Noticed		Detected	Number detected per thousand surveys
0.10	0.10	0.25		0.0025	2.5

The figure below displays the same information graphically, by showing how the chance to detect an incursion changes as the chance that an incursion is actually present changes.

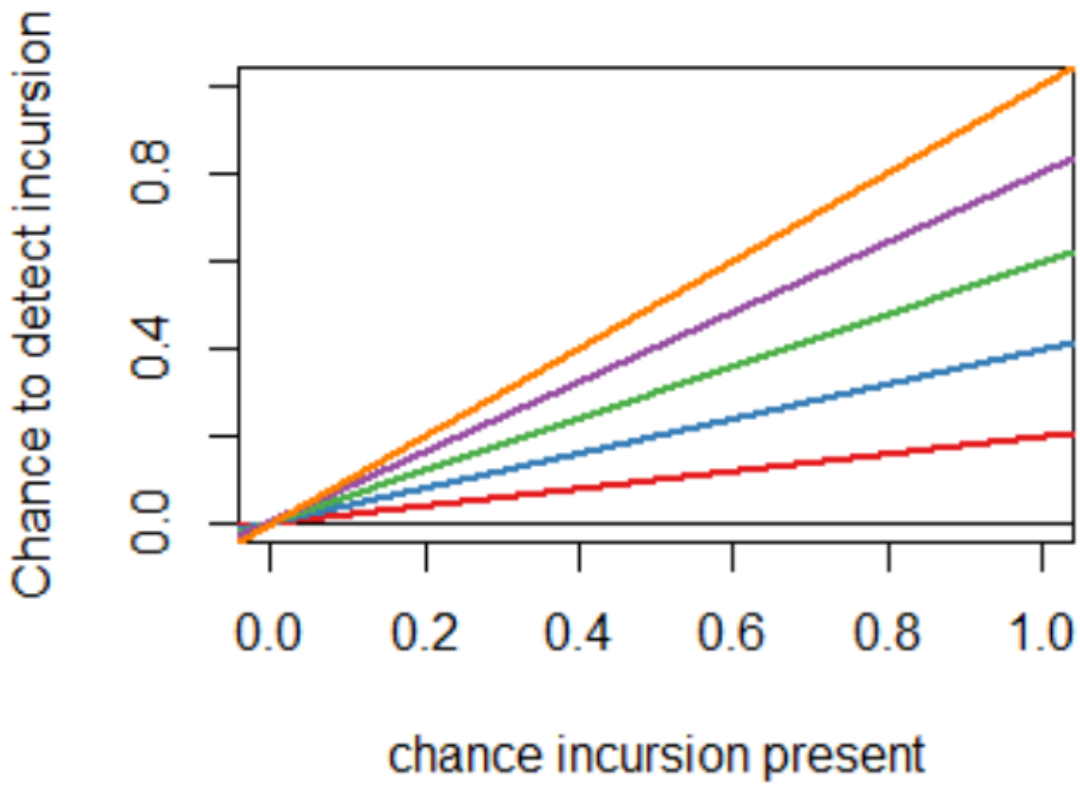


Figure 1: plot of the chance to detect an incursion versus the chance the incursion is actually present. The orange line shows the best-case scenario where the entire glass house is surveyed each time and the incursion is detected every time it is present. Other lines show poorer scenarios, such as when 80 % of the glasshouse is surveyed (purple), 60 % (green), 40 % (blue), or 20 % (red). All lines in this plot assumes that the incursion is detected when the correct portion of the glasshouse is surveyed. When the incursion is likely to be missed, surveys are less likely to detect them.

Appendix 1: additional background analyses for section 1

Summary:

I (William Godsoe) was asked to evaluate sampling protocols for detecting whiteflies in tomato growing facilities. Whitefly growers sample sites throughout each facility and score plants based on the level of whitefly infestation. These scores are commonly between 0 and 6, though higher numbers are sometimes noted. I submitted a preliminary report in November, using simulations in the computer program R to develop a model for the distribution of whitefly scores and asking how likely it would be to find samples exceeding thresholds used by growers (parts 2-9 of this report). Feedback from that report suggested that growers would be interested in an approach, which was as simple as possible to evaluate the effectiveness of sampling. In response, I have added a new section, (part 1) which provides a simple formula to determine how reliable a sample is likely to be, based on the size of a sample across the glasshouse and the score, which is expected. I have provided illustrations of this formula and an excel spreadsheet which automates the calculations. The conclusion of this section is that increasing sample size in a region of interest is quite valuable up to several dozen samples, after this point, increasing samples still has a benefit but the benefit is weaker.

Part 1: Formula for how many samples will be needed to compute average scores

There seems to be a desire to have a simple rule of thumb for how many samples are needed in the glasshouse. Standard statistics suggests an approach to get one such formula. Let us call the real average score across all bays in the glasshouse *X*, for example, in week 45 of 2020 the average score across all samples in the glasshouse was 4.6. When we take a sample of size n, we can get an average score from the sample, but it is likely to be off from the correct average. To find how far the average score is from the true score, we can use a formula for confidence intervals. Approximately 95% of the samples you take will be bigger than:

$$X - 4.3 \times \sqrt{\frac{X}{n}}$$

and smaller than:

$$S + 4.3 \times \sqrt{\frac{X}{n}}$$

In words, these formulae state that when the true average score is *X*, we expect most samples to be in a range of values which starts at *X* minus 4.3 times the square root of *X* divided by the sample size and ends at *X* plus 4.3 times the square root of *X* divided by the sample size. When this range of values is big, it won't be clear what whitefly score in the glasshouse is. When this range of values is small, we will be quite clear what the average score across the glasshouse is. The question then is when adding additional samples will make the range of possible values substantially smaller.

When the sample size is small, there is a big benefit in collecting more samples. So increasing your sample size from 5 locations to 10 locations offers a substantial improvement. As the sample size gets bigger, the benefits of more samples get weaker. So increasing your sample size from 105 locations to 110 locations does very little to improve your understanding of the outbreak. Below I provide plots to illustrate this effect. To make it more accessible to growers, I have created excel spreadsheet which will do the calculation automatically (Figure 1). This calculation could also be done with a hand calculator, I would be happy to provide examples upon request. Note that this formula is just a rule of thumb, It will not be exactly right.

Sample Siz	Average S	tvalue	Lower valu	Upper value
5	4	4.3	0.15	7.85
10	4		1.28	6.72
105	4		3.16	4.84
110	4		3.18	4.82

Figure 1: Spreadsheet with four example calculations for sample sizes 5,10, 105 and 110 respectively. This spreadsheet (titled checkConfidenceIntervals.xlsx) will be attached with the report.

Details of calculation

To obtain the formula above, I assumed that the scores approximately follow a Poisson distribution. This means that the average score is approximately equal to the variance in scores. I then assumed that the number of samples collected was large enough to be approximately normally distributed. This means that we can calculate 95 % confidence intervals using a t-distribution. These confidence intervals get wider when sample size is small. So to be conservative and so I used the confidence intervals for a t-distribution with two degrees of freedom.

Evaluation

One way to test this formula is to take the entire Reporoa dataset (with 180 samples per week), then artificially create small samples of size *n*. We can get the computer to create many such artificial samples, and then ask how different the average of these artificial samples are from the correct value. In the plots below (Figure 2 and 3), we have done this for different sample sizes from 1 up to 180. For two weeks (black lines). When only one or two samples are collected, the average score varies a great deal. As we increase the sample size from 1 to about 20 the samples get far closer to the correct value, after about 20, increasing the number of samples helps, but not nearly as much, by the time the sample size is up to 180, we know exactly what the score is.

The formula I have suggested above, gives an idea of this effect. Initially, increasing the sample size dramatically improves our understanding of the average score, but past about 20 points, adding more samples improves our understanding slowly. Note that the formula is a bit too wide when sample size is higher than about 100.

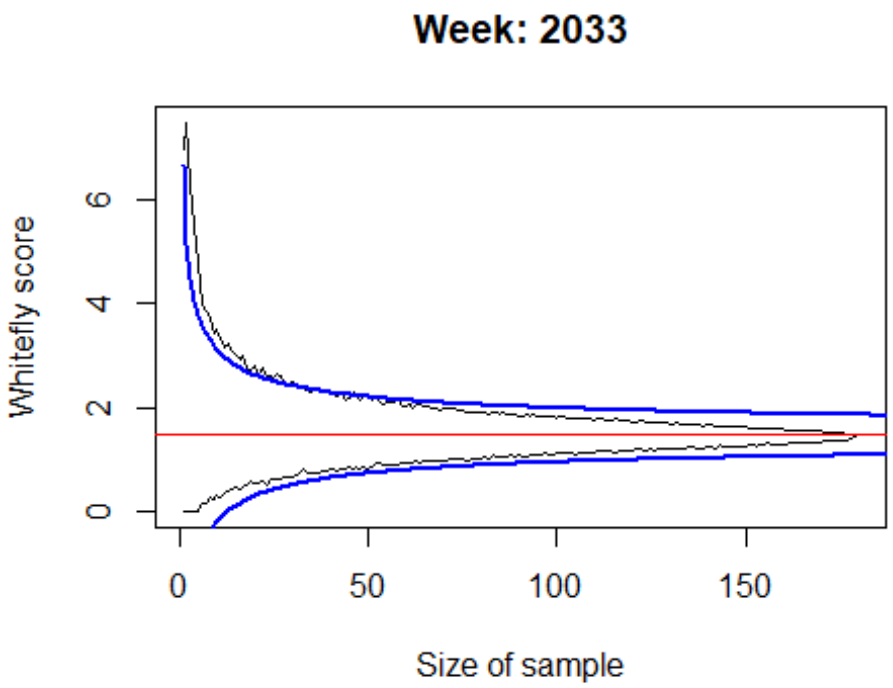


Figure 2: Plot of estimates of whitefly score versus the correct score for week 33 of 2020 in the Reporoa facility. In this week, whiteflies are uncommon, with the average score across the facility of a bit less than 5 (horizontal red line). The black line indicates the range of values observed by creating samples of sizes from n=1 to n=180. When the sample size is small the average scores varies a great deal, but as the sample size increases up to 180, virtually all of the samples are close to the average across the facility. To calculate the black line I had to make my own code in R, which will be hard for others to use. To get approximately the right answer we can use the formula I suggested above which can be calculated in excel.

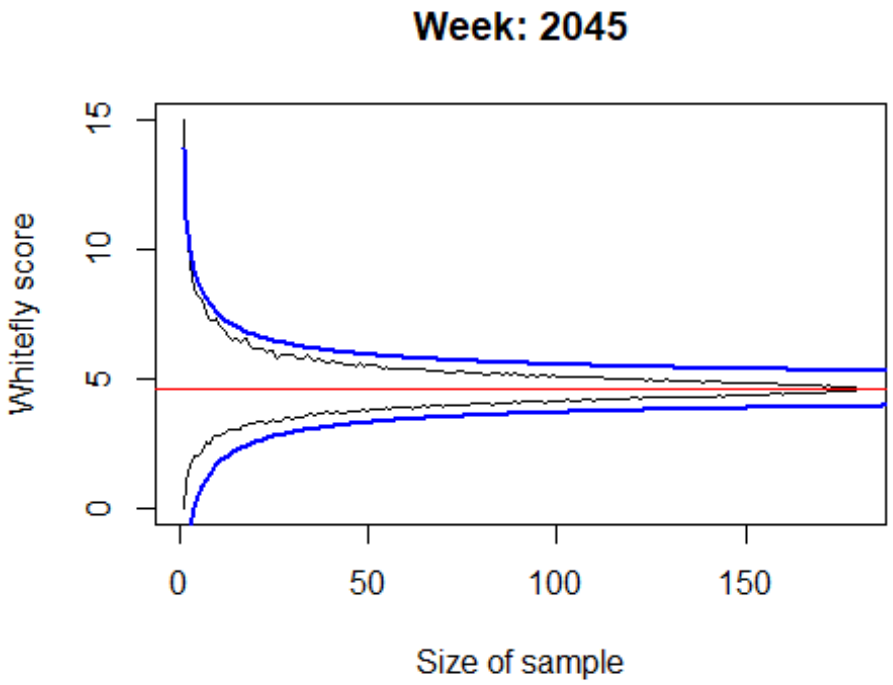


Figure 3: Plot of estimates of whitefly score versus the correct score for week 45 of 2020 in the Reporoa facility. In this week whiteflies are abundant and the average score across the facility is a bit less than 5 (horizontal red line). The black line indicates the range of values observed by creating samples of sizes from n=1 to n=180. When the sample size is small the average scores varies a great deal, but as the sample size increases up to 180, virtually all of the samples are close to the average across the facility. To calculate the black line I had to make my own code in R, which will be hard for others to use. To get approximately the right answer we can use the formula I suggested above which can be calculated in excel.

Part 2: an outbreak where flies are distributed at random

In part 1 I described a tools to evaluate the usefulness of samples to find average scores. Standard statistics formulae tools work well for averages, but in our initial discussion there was an interest in finding how many samples exceed a threshold. Therefore, in what follows I describe increasingly more realistic models to evaluate when samples are likely to exceed thresholds that we were told would be of interest to growers.

Below is my first simulation. This is of a glasshouse where the average number of insects per plant is 2, but the number of insects detected on each plant can vary due to luck (i.e. randomness). I assume that the glasshouse has 3000 sections (i.e. 300 rows, each of with 10 sections per row). In each section I simulate a sample of three plants. I then take an average number of insects per plant.

This simulation shows us how much variability to expect in samples across the glasshouse. Figure 1 shows what to expect across all samples.

I use this plot as a tool to picture decisions in the glasshouse. For example, we had discussed what it means for the number of whiteflies in a sample to be 2 or greater. This plot shows that when the average number of flies is 2, many samples still have less than 2 whiteflies (Figure 4).

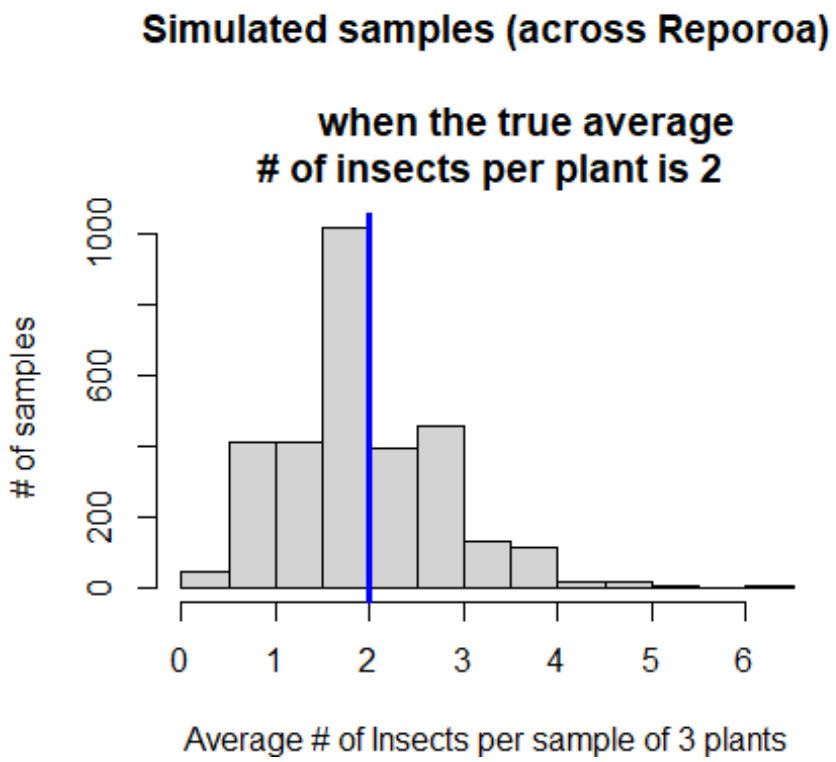


Figure 4. Distribution of whiteflies across a simulated glasshouse. The blue line a threshold at 2, we can see that roughly half the samples have fewer whiteflies than 2.

Part 3: simulations of different thresholds when whiteflies are random

In practice, we don't know what the true number of whiteflies per plant is. Below I simulate outbreaks, where the average number of whiteflies per plant across the glasshouse varies from 0 to 20. I then ask how many samples exceed different potential thresholds: 2, 5, 10. This shows that when the number of flies per glasshouse is high enough, we find enough whiteflies to exceed thresholds in each of the samples (Figure 5).

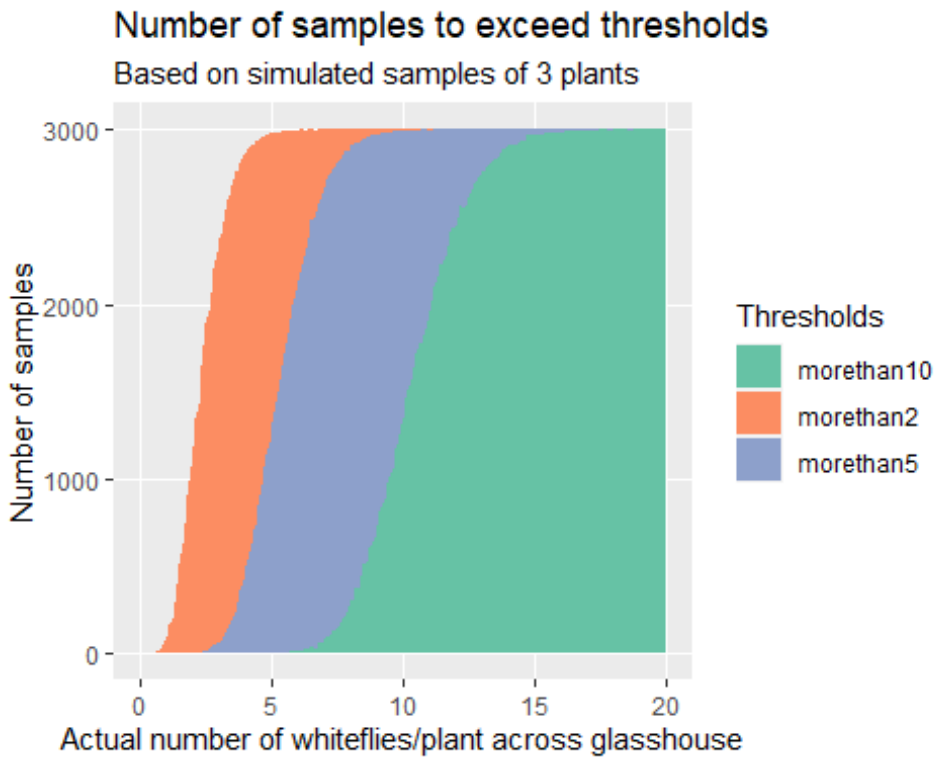


Figure 5. Plot of the number of samples that exceed different detection thresholds in simulated glass houses where the mean number of samples per plant varies from 0 to 20.

Part 4: simulations when whitefly outbreaks are spatially clumped

Whitefly outbreaks are not likely to be evenly distributed across the glasshouse. Instead, we'd expect to see big outbreaks in some sections and few whiteflies in other sections. Below we show how to model an outbreak which is spatially clumped. We'll need to discuss how best to link this model to glasshouses.

The model we showed above can be modified to include a measurement of how clumped we'd expect whiteflies to be. Figure 3 shows one simulated glasshouse with spatial clumping. When clumping is included the some samples have 0 flies and other samples have a large number of flies. Compared to the flies distributed at random in figure 5, this this clumping can make it harder to detect outbreaks.

The clumping is controlled by a number called "overdispersion". Below is a previous version of the model I wrote based on a worst case scenario where overdispersion was quite high. You'll see below i've used data in Reporoa to check this model.

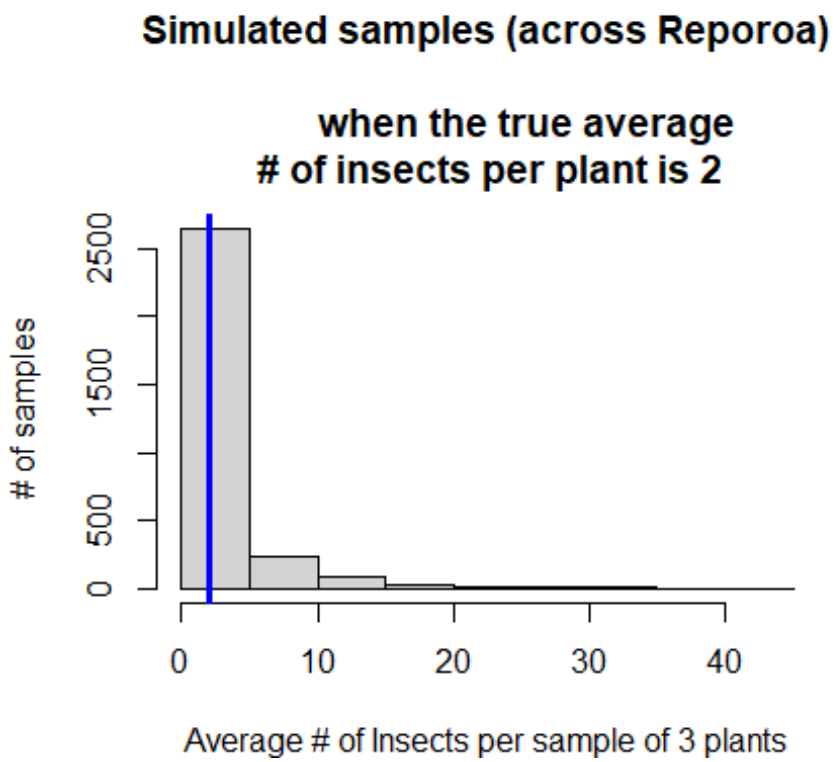


Figure 6. Distribution of whiteflies across a simulated glasshouse. The blue line a threshold at 2. We can see that some samples have many flies, while other samples have very few flies.

Part 5: simulations of different thresholds when flies are clumped

Below is a summary of simulations where flies are clumped (Figure 7). I assume that the average number of flies per glasshouse varies from 0 to 20. With this simulation, we can see that many samples across the glasshouse remain below the detection thresholds. This means that when the distribution of insects is clumped, it will be harder to use thresholds to detect the outbreak. Our next steps are understand how much of a problem this clumping will be.

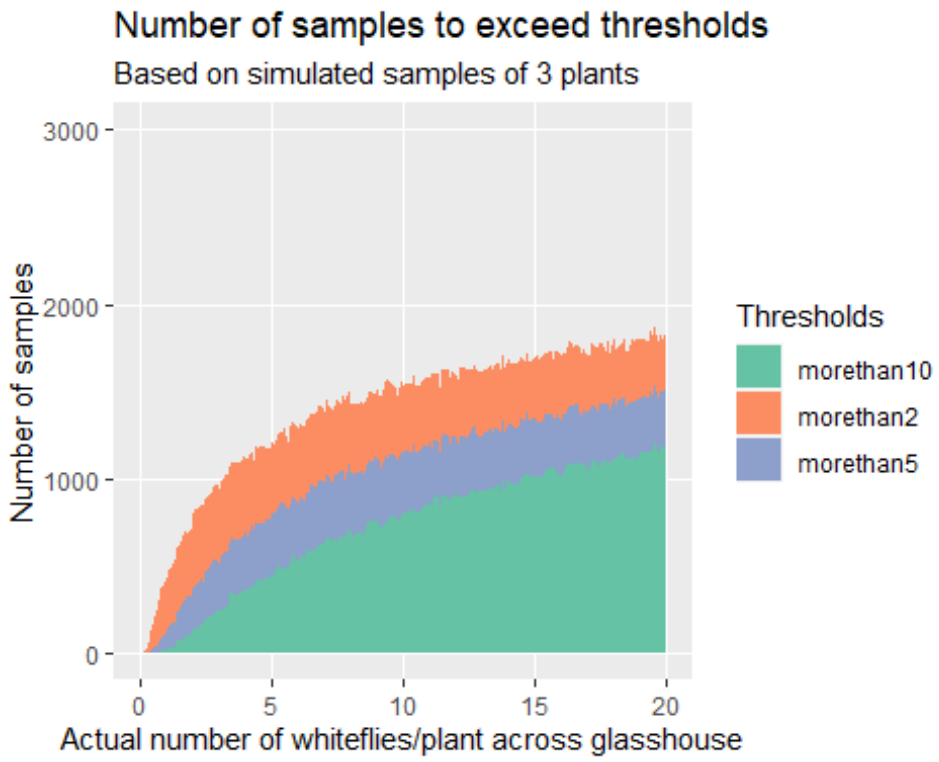


Figure 7. Plot of the number of samples that exceed different detection thresholds in simulated glass houses where the mean number of samples per plant varies from 0 to 20, and flies are spatially clumped.

Part 6: Testing if early whitefly outbreaks are spatially clumped

To test whether the simplified model described in parts 2-5 is realistic I've re-formatted the Reporoa dataset to test for clumping in the distribution of flies (Figure 8).

```
##
## -- Column specification --------------------------------------------
--------
## cols(
##   Row = col_character(),
##   ID = col_double(),
##   score = col_double(),
##   Week = col_character(),
##   Section = col_character()
## )
```

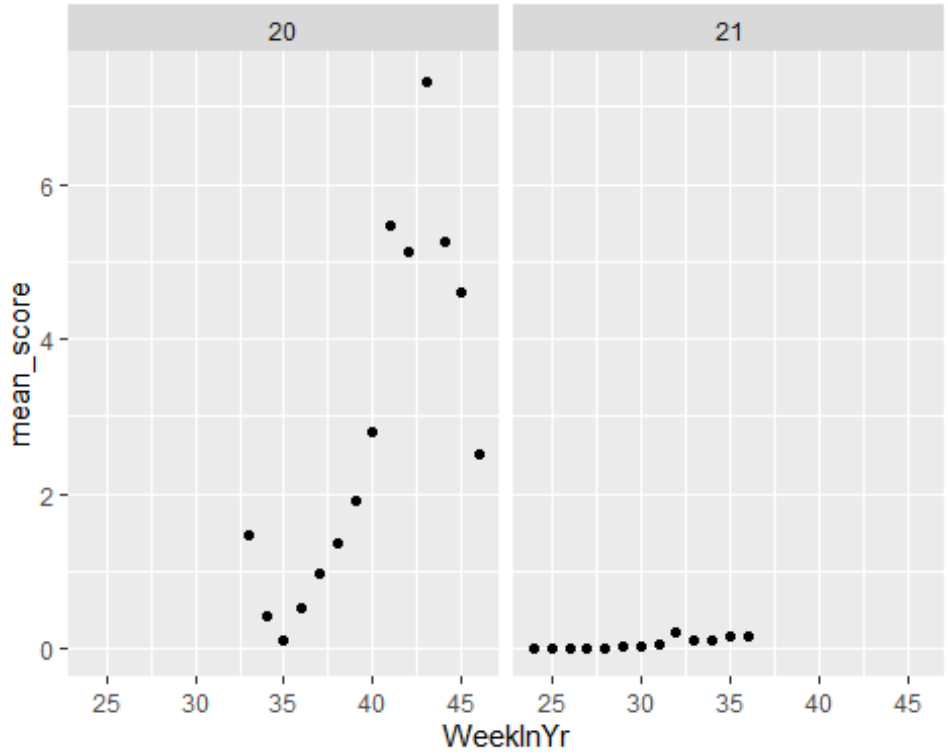


Figure 8: the average score of whiteflies for both 2020 and 2021. We can see a big outbreak in 2020 and a few flies in 2021.

Part 7: Examples of clumping when scores are large

Below you can see the number of samples with different score values for a given week (Figure 9). These plots allow us to compare the actual distribution of scores (black lines) with what I assumed the scores would be for the simple model with no clumping (blue lines). The figure on the left shows the results for week 2033 when the outbreak is starting to grow at Reporoa. The figure on the right shows week 2045 where the outbreak is quite large.

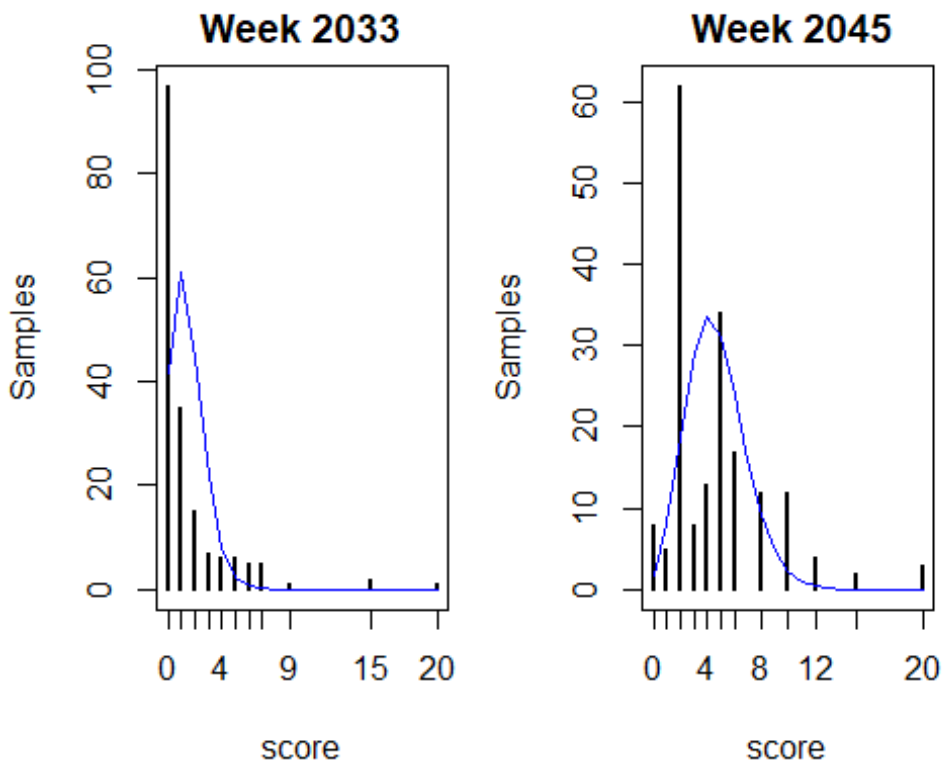


Figure 9: the distribution of scores for samples collected in weeks 33 and 45 of 2020 (left and right respectively). The black lines show the observed number of samples with each score while the blue lines show what we would expect if the counts came from a poisson distribution.

Part 8: Scores are close to random when the number of flies is small

When there is no clumping, we expect variability in scores to increase about as fast as the average score. This suggests a simple check, plot the variability in scores versus the average score for each week in Reoporoa (Figure 10). If there is no clumping the scores should roughly line up.

A few technical notes for the math inclined. Technically i'm testing the assumption that the mean is equal to the variance. This assumption is made for the Poisson distribution. This test is a bit crude, but it makes a plot rather than a complex stats analysis. Even before the analysis it was pretty clear that the Poisson distribution will likely be unrealistic during the middle of the outbreak. This is because the Poisson distribution works with whole numbers, but the scores are qualitative.

```
par(mfrow=c(1,2), mar=c(4,4,2,1))
plot(var_score~mean_score,
     RepByWeek,
     pch=16,
     xlim=c(0,0.6),
     ylim=c(0,1.1),
     xlab="mean score",
     ylab="variance in score",
     main="Data w. small scores")
```

```
abline(0,1, col="blue")
text(RepByWeek$var_score~RepByWeek$mean_score,
     labels=as.character(RepByWeek$Week), pos=4, cex=0.5)

plot(var_score~mean_score,RepByWeek, pch=16, main="All the data")
abline(0,1, col="blue")
```

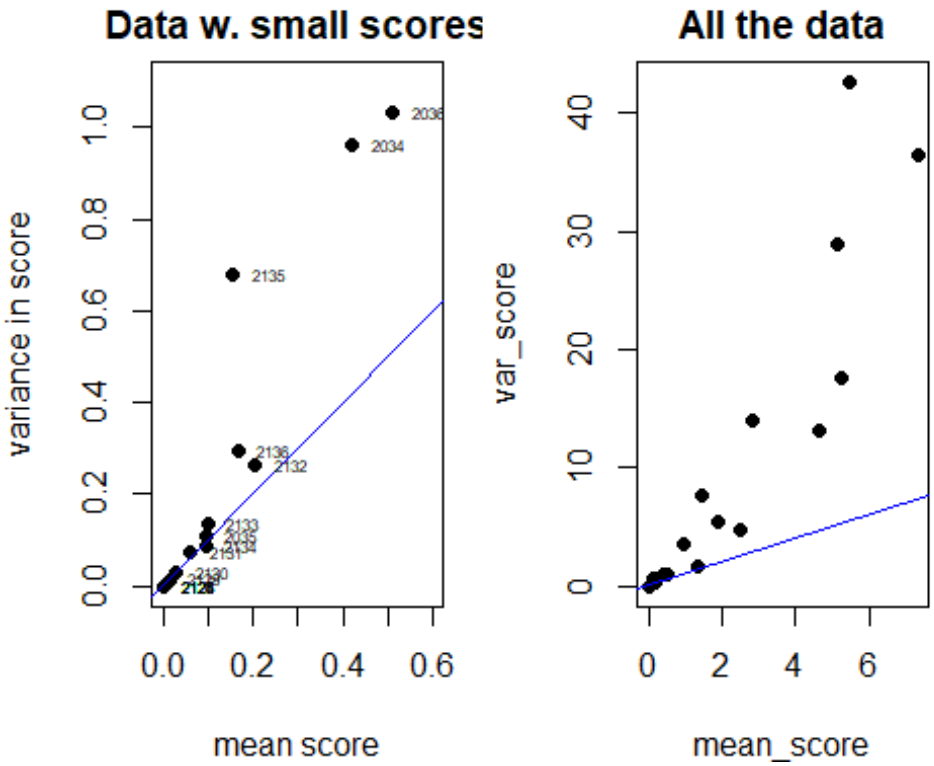


Figure 10: Plot of the variance in scores versus the mean scors across weeks in Reporoa. The blue line shows the relationship we would expect if scores were Poisson distributed. In weeks where scores are large, the variance is higher than we would expect under a poisson distribution.

Therefore the randomness assumption is close to ok when average scores are low across the glasshouse, but better models are needed when scores are high.

Using just the weeks where counts are close to zero, suggests that the Poisson model is a good fit to the data. But this is really an informal check and too optimistic, i took the portion of the data that looked nice and ran a model on it. Even this estimates that the variance increases twice as fast as we'd expect.

```
summary(lm(var_score~mean_score,RepByWeek[RepByWeek$mean_score<0.6,]))

##
## Call:
## lm(formula = var_score ~ mean_score, data = RepByWeek[RepByWeek$mean_sco
re <
##     0.6, ])
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -0.15776 -0.04867 -0.00636  0.01840  0.36430
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.01840    0.03687  -0.499    0.625
## mean_score   2.14511    0.19601  10.944 3.02e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1162 on 14 degrees of freedom
## Multiple R-squared:  0.8953, Adjusted R-squared:  0.8879
## F-statistic: 119.8 on 1 and 14 DF,  p-value: 3.023e-08
```

Part 9: Analysis including data from reporoa

The data from Reporoa shows that when scores are extremely low, there is no evidence of clumping. This was the case for week 2132. When scores are a bit higher, the variance is about twice what we would expect. This is the case on week 2034. I've modified the model for this second, possibility (Figure 11).

Technical note, I used the overdispersion parameterization of the negative binomial distribution in R. I've revised the code so that that the variance is equal to twice the mean.

```
## Warning in rnbinom(dim(reporoa)[1], mu = meanWhitefly, size = s): NAs pr
oduced

## Warning: Removed 1 rows containing missing values (position_stack).

## Warning: Removed 1 rows containing missing values (position_stack).

## Warning: Removed 1 rows containing missing values (position_stack).
```

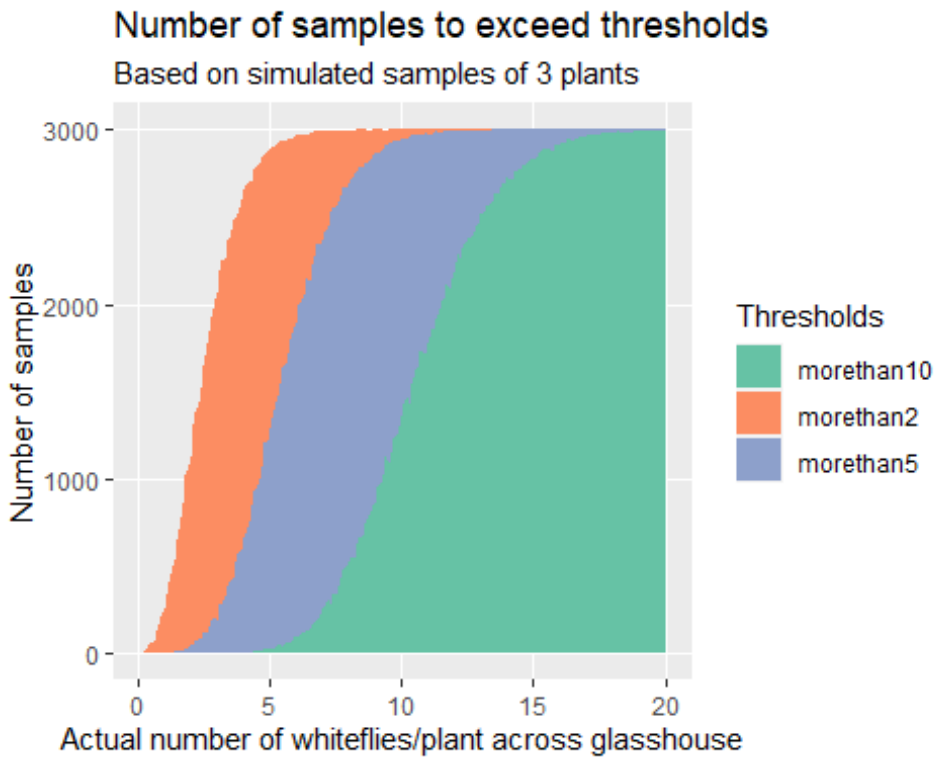


Figure 11: Modified plot of the number of samples to exceed thresholds using data from Reporoa to adjust the relationship between variance in scores and the mean scores. With this correction, the plot looks only slightly dissimilar form the original analysis.