Gold3 - the use of regression analysis

Grower Fact Sheet
#12
RELEASE DATE:
April 2019

Introduction

Consumers consistently indicate that Taste is among the most important quality attributes of Zespri kiwifruit. To that end, Growers are encouraged to meet certain preference thresholds for all commercial varieties and are incentivised via a payment mechanism to produce tasty fruit.

Consumer research has shown that dry matter content, the percentage of a fruit's fresh weight remaining once all the water has been removed, is strongly linked to consumer liking (higher dry matter = greater preference). The advantage with dry matter measurements is that they can be made at or before the fruit are harvested and used to infer possible consumer responses to the fruit after they have been stored and ripened many months later. Thus, information on the dry matter levels of fruit not only provides a useful decision-making tool for the Grower in deciding when to harvest, but it also is a useful decision-making tool for Zespri in managing distribution and marketing of kiwifruit.

Zespri's OPC team have produced a suite of Need to Know resources to help Growers make the harvest decision, and to ensure that you're well prepared when the day arrives. You'll find them all in the Need to Know library, and under the various topics in the Maturity and Harvest page on Canopy. There is also a direct link to pertinent resources in the 5 March 2019 KiwiTips special edition.

Purpose of this factsheet

In the 2016 harvest season, Zespri introduced taste by size for Gold3, where a regression (trendline) is developed based on the 90-random and 60-small fruit sample and is used to assign a taste (TZG) value for each fruit size. This is used in combination with your actual fruit size profile at packing to calculate your taste payment.

The purpose of this factsheet is to supplement the OPC resources provided by primarily addressing two frequent questions raised by Growers to NZKGI:

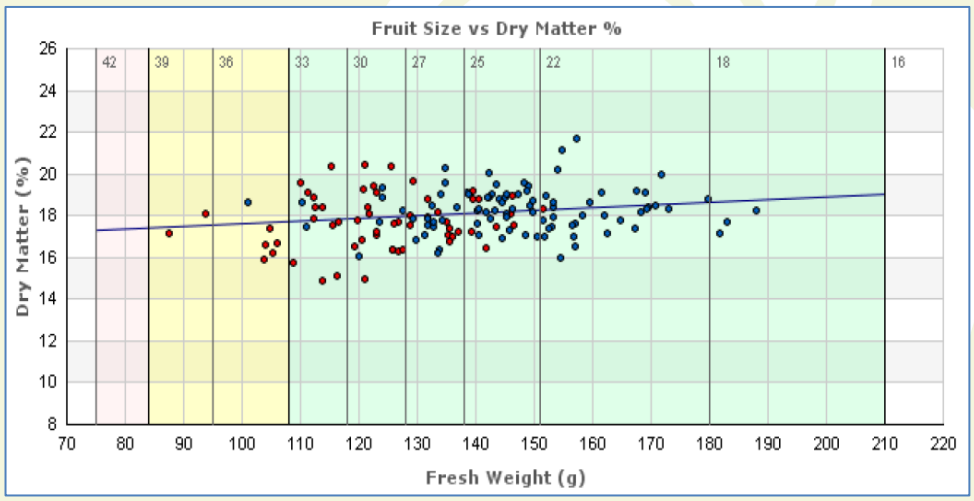
1. What is regression?
2. Why are small sized Gold3 fruit pieces included in the Gold3 Taste calculations?

1. What is regression?

One of the things that we know about nature is that there are often quite simple relationships between the fruit and its growing conditions. In Gold 3 we have a variety which shows a strong influence of fruit size on several quality variables. Regression is a statistical technique used to describe relationships among these variables. Regression explains how one variable (referred to as the dependent variable, on the Y-axis) may be related to other variables (referred to as the independent variables, on the X-axis). Put another way, the dependent variable (Y) is the main factor that we're trying to understand, and the independent variables (X) are the factors that we hypothesize have an impact on the dependent variable. The purpose of regression is to try to find the best fit line (trendline) or equation that expresses the relationship between (Y) and (X).

In the case of Gold3 we know that the quality variables of dry matter, colour and brix all depend on the size of the fruit (i.e. the size of the fruit tells us something of the likely relative values of the other 3 quality variables). The way that regression actually works is it takes all the data available for the dependent and independent variables and calculates the line that fits the data with the smallest error. As part of this calculation the regression determines how well the data "fits" (called the R^2). The closer the data to the regression line, the better the R^2.

Referring to the diagram on the right, taste for Gold3 is based on a 90-fruit random maturity area sample (the blue dots), and an additional 60-fruit small-size targeted sample (red dots). The regression line drawn through the sample, along with the variability of the sample, is used to calculate taste by size.



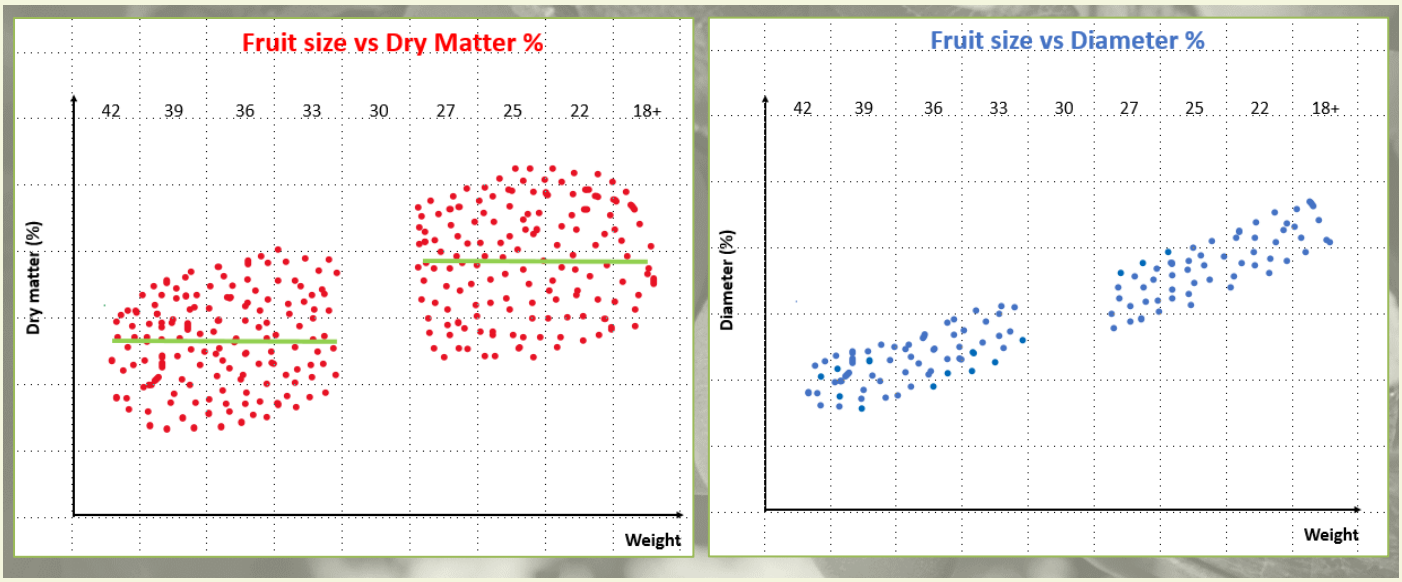
We know that the larger fruit generally has higher average dry matter than the smaller fruit. If we were to use the overall average for all fruit sizes, we would be tending to over-estimate the dry matter of the smaller fruit sizes and under-

estimate the dry matter of the larger fruit sizes. The regression approach overcomes this problem, giving better dry matter estimates across all fruit sizes. This is discussed in more detail in the following section.

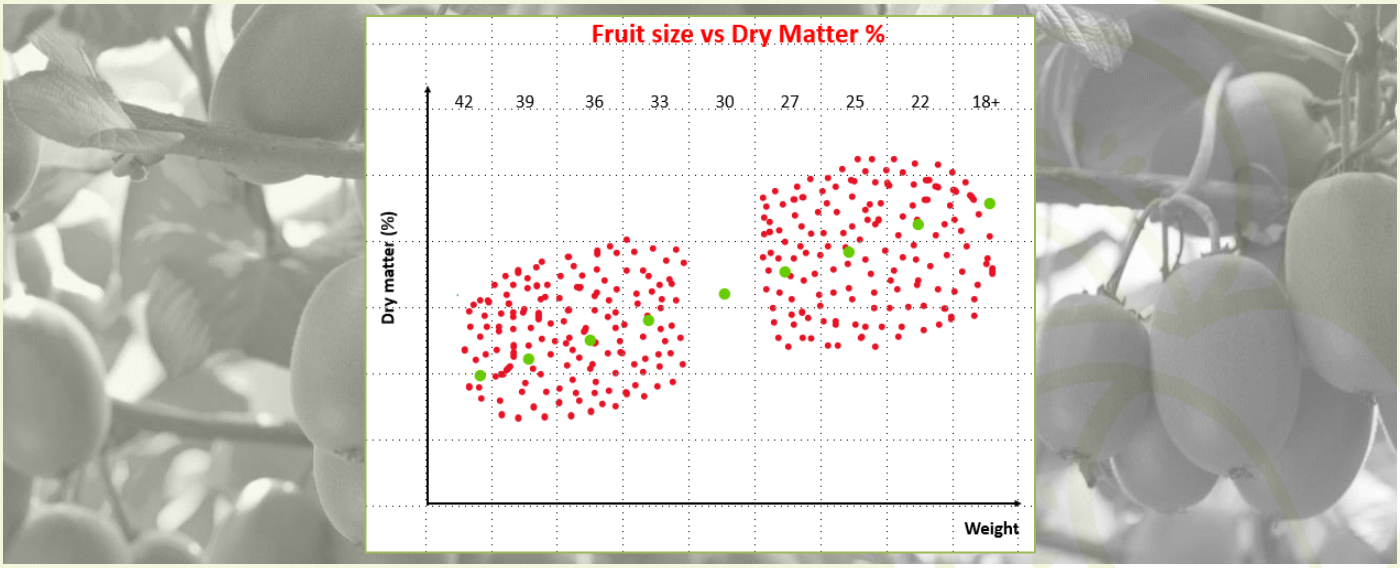
2. Why are small sized fruit pieces included in the Gold3 taste calculations?

The Industry Supply Group's (ISG) Maturity Review sub-group met with statisticians late last year to review whether small sized fruit pieces should be included in the Gold3 taste calculations. Keeping in mind that there is a size to dry matter relationship in Gold3 where the larger fruit in a maturity area are usually higher dry matter than the smaller fruit, let's look at the following illustration.

Suppose two 90-fruit samples are taken from a maturity area. One sample consists of all fruit below size 30 and the other sample consists of all fruit above size 30. Weight, dry matter and diameter are measured at the lab. We might get a pattern like the one below for each of dry matter and diameter. The dry matter pattern is quite variable (i.e. the red dots are widely spread) but the diameter pattern (blue dots) is more concentrated and much less spread out. If we did an eyeball estimate of the average of the samples, we might draw a line as indicated with the green line in the dry matter graph below.

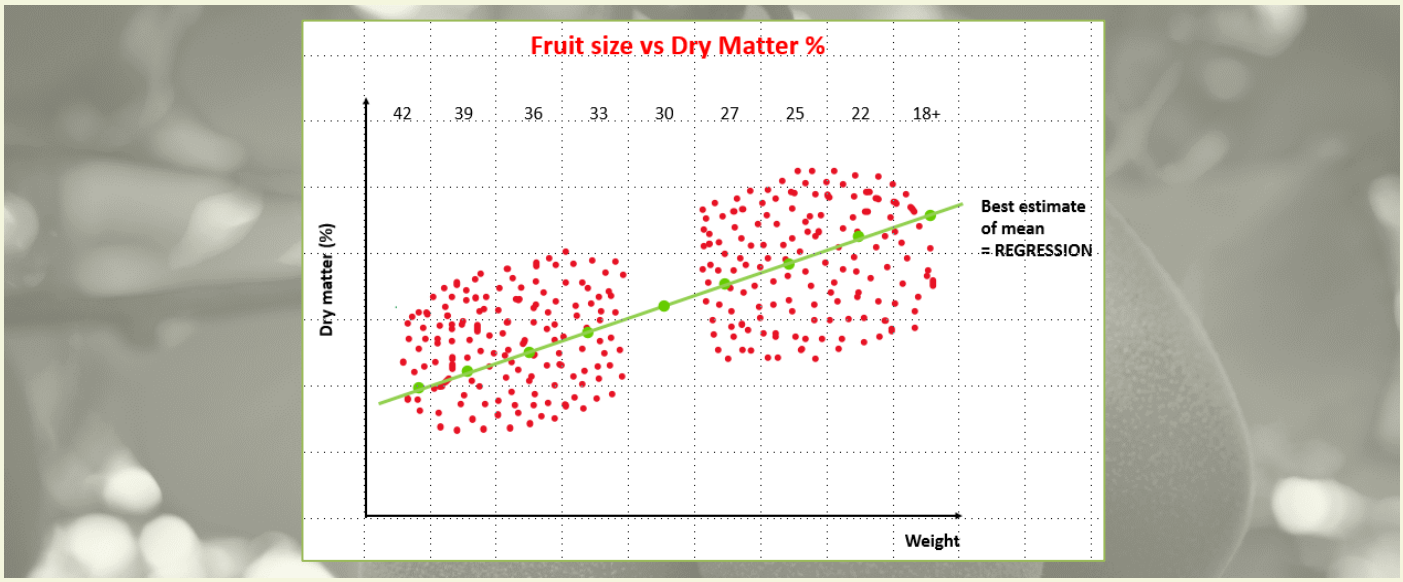


Note in the graphs above that there is no fruit in the count 30, so how is the average dry matter for the 30s estimated? This is where regression is helpful; the green dots below show the regression's "best estimate of the mean (average)" for each count size.



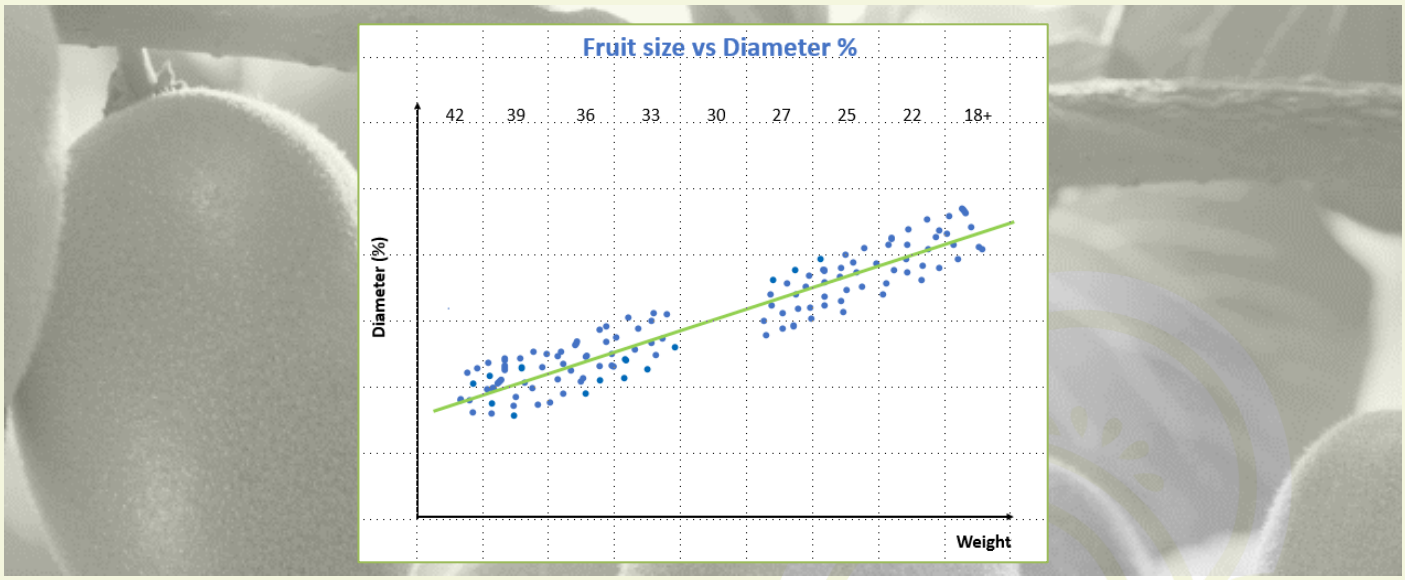
In other words, the regression is able to complete any missing estimates (e.g. the 30s) by using the rest of the sample data. Regression is used instead of using the individual fruit pieces because some sizes may have very few fruit pieces sampled, but we still want to estimate the average of each size e.g. there could be very few pieces in sizes 39 and 33.

The regression draws a line through its estimate of the average dry matter for any weight (or any size band) i.e. through the green dots).



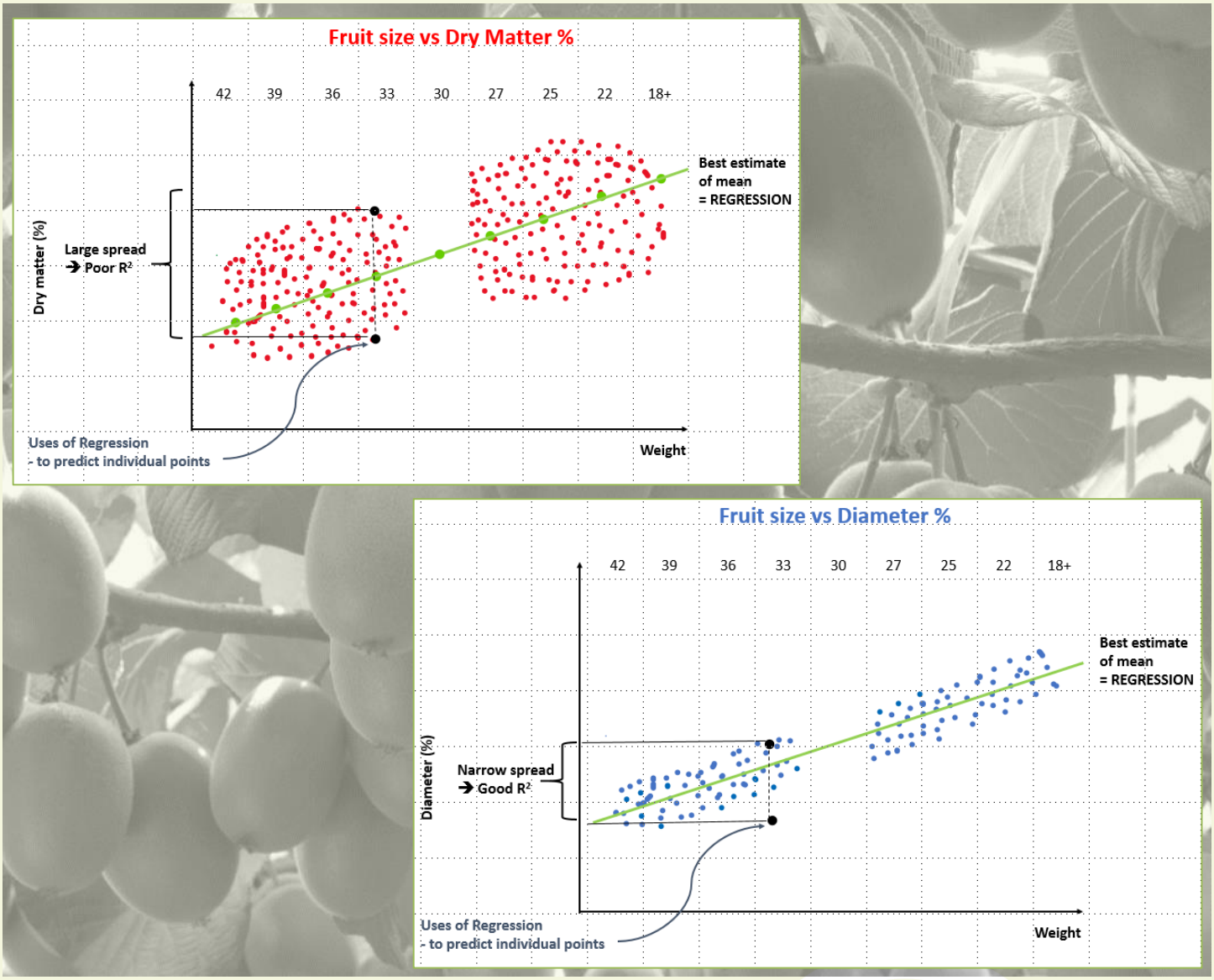
The regression also has a calculation to demonstrate how close the sample data is to the line and therefore how well the line "fits" the data (R^2). We cannot expect to get a very good fit for the dry matter to size relationship in this example because the dry matter is so variable.

If we now think about the fruit diameter data, the regression line better fits the data (green line) because the data is significantly less spread out.



We can therefore conclude in this example that both dry matter and diameter have a relationship with size but the regression line is weak for the dry matter correlation and much better for the diameter correlation because the data is less spread out i.e. regression has enabled us to predict a value for something (e.g. diameter) by measuring something else (e.g. weight). If we measure the weight of a fruit and want to predict the diameter of it then the regression equation will estimate the diameter based on the green line.

If we try to do the same thing with the prediction of dry matter for a given fruit weight then there is a lot more noise. However, we don't want to predict dry matter for individual fruit; we want to predict the average dry matter across all fruit (i.e. points along the regression line). The regression needs fruit pieces across the entire size profile to accurately predict the average dry matter of each size, so even if the smaller fruit are not being harvested they are still required to help estimate average dry matter of all sizes.



ISG Maturity Review sub-group

After extensive discussion and consideration by the ISG Maturity Review sub-group, it was evident that increasing the range of fruit weight in the sample helped improve the accuracy of the estimates of dry matter by fruit size in Gold3 for every fruit size. There is a size to dry matter relationship in Gold3 where the larger fruit in an orchard generally has higher average dry matter than the smaller fruit. Sometimes this is a very obvious regression slope and sometimes it is relatively flat. Other times the fruit on the orchard are so variable that it is hard to deduce anything. The regression line tends to be sensitive at the extremities. Including small fruit tends to steepen the slope of the line, so the estimated dry matter for larger sized fruit is increased. Taking out the small fruit can flatten the line, whereby the estimated dry matter for smaller sizes increases as the estimated dry matter for larger sizes decreases.

Conclusion

The Gold3 taste by size calculations use fruit weight to get better predictions of dry matter for each size class. Inclusion of the smaller fruit sizes (even if they are not being harvested) improves the prediction of dry matter for all fruit sizes.