

## Questions

### Section 1: Phenotyping Corn Seedlings.

Q1. What is quantitative trait? Is leaf angle a quantitative trait?

Q2. Which maize genotype that was phenotyped in your class had the largest 2<sup>nd</sup> leaf angle? Which one had the smallest, most upright 2<sup>nd</sup> leaf angle?

**Standard deviation** (SD) is an important statistic that is used to measure the amount of variation in a set of values. SD can be calculated in Microsoft Excel with the function *STDEV()* by typing = *STDEV()* in a cell and then highlighting the numbers that you want to measure SD for.

Q3. What is the SD of the 2<sup>nd</sup> leaf angles within each genotype that you phenotyped? What is the SD of the 2<sup>nd</sup> leaf angles across all genotypes that were phenotyped in your class? What do these statistics say about the data you collected?

### Section 2: Phenotype leaf angles of corn seedlings using computational methods.

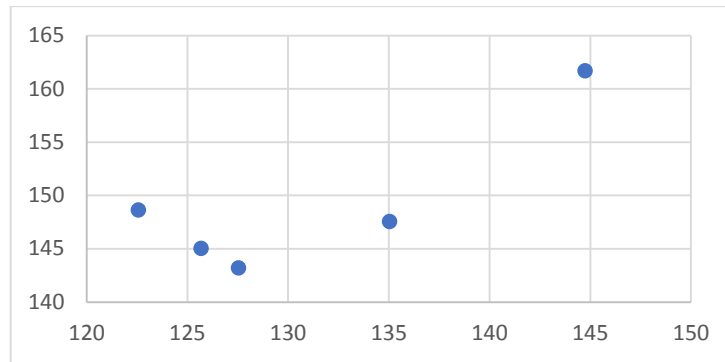
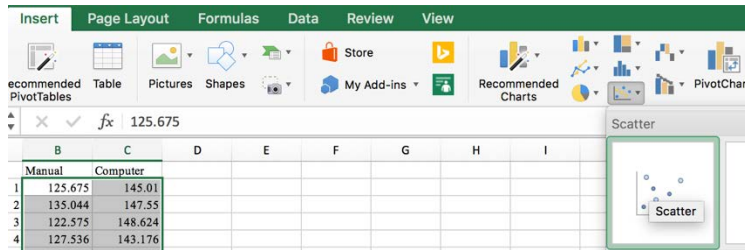
Using both manual and 2D image-based (provided by *PlantCV*) data that you collected on the leaf angles of the 2<sup>nd</sup> leaves, you can test whether the image-based data are good proxies for the “ground truth” hand measurements using linear regression.

How to do it...

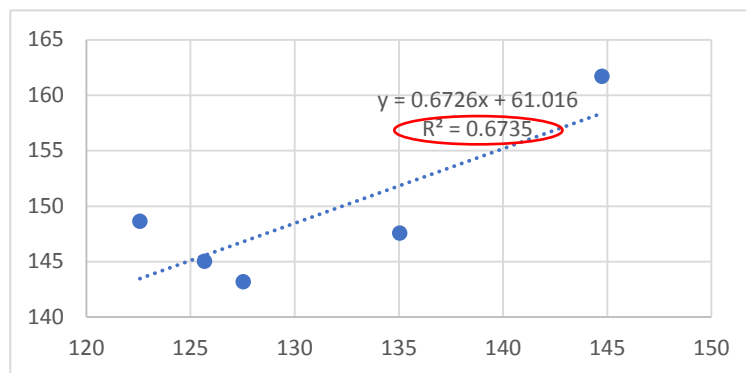
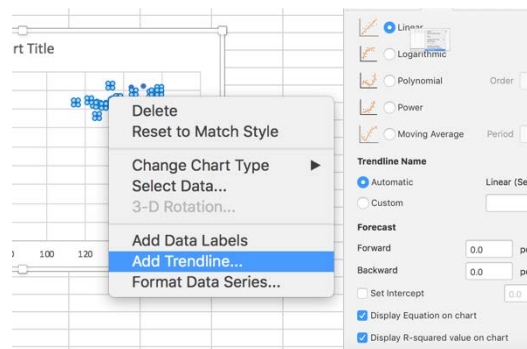
- Input your data from both manual and image-based measurements into Microsoft Excel (see diagram to the right for a template on how to input your data into Excel). Excel can be used to model the linear relationship between the two measurements.
- Highlight the data.
- From the “Insert” tab, click on the Scatter Plot.

A	B	C	D
Plant	Manual	Computer	
1	125.675	145.01	
2	135.044	147.55	
3	122.575	148.624	
4	127.536	143.176	
5	144.752	161.685	

This will generate a Scatter Plot which plots the Manual data on the x axis and the image-based data on the y axis.



- d. Select any data point on the Scatter Plot and right click to open the dropdown menu. Choose “Add Trendline” and then click “A linear relationship”, “Automatic”, “Display Equation on chart”, and “Display R-squared value on chart”. A trendline will be added on your chart over the data points. In addition, the  $R^2$  value, the indicator of the goodness of fit (shown in the red oval), and the equation of the trendline ( $y = mx + b$ ) will be shown on the Scatter Plot.



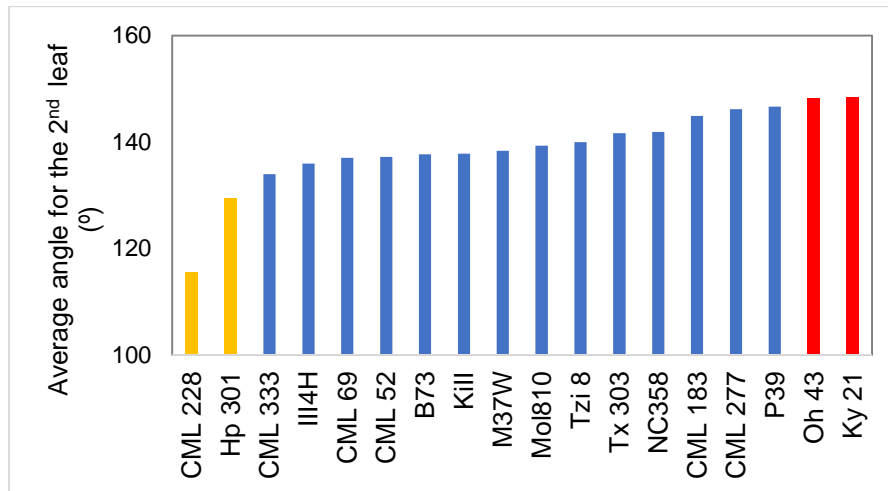
y is the response variable (image-based measurement);  
x is the explanatory variable (hand measurement);  
m is the linear slope related x to y;  
b is the intercept.

How do you interpret the trendline and the  $R^2$  value in your scatterplot? What do they mean?  
The trendline is generated with the smallest possible distance between the line and all data points. The  $R^2$  value is the Coefficient of Determination, and describes how closely the data points are to the trendline. An  $R^2$  value of 1 means there is a strong relationship while an  $R^2$  value of 0 means there is no relationship at all.

Q4. What is the  $R^2$  value of your data? How well do the computational measurements fit the hand measurements?

### Section 3: Genotyping SNPs

The graph below displays the average values of 2<sup>nd</sup> leaf angle measurements from seedlings of 18 diverse corn genotypes. Among the 18 genotypes, Oh43 and Ky21 have large 2<sup>nd</sup> leaf angles, while CML228 and Hp301 have small ones.



The table beneath shows the 2<sup>nd</sup> leaf angles of 4 individuals from each of the 4 genotypes, and SNPs at 10 different positions in their genome sequences. To determine whether the SNPs in the table are correlated with the 2<sup>nd</sup> leaf angles, compare the SNPs of plants with large (red) and small (yellow) leaf angles, search for patterns in the data, and answer the questions.

**Table 1. SNP markers at 10 different loci in corn genotypes with large (red) and small (yellow) 2<sup>nd</sup> leaf angles.**

SeedlingID	2nd leaf Angle (°)	Marker1 Chr2	Marker2 Chr6	Marker3 Chr1	Marker4 Chr5	Marker5 Chr5	Marker6 Chr8	Marker7 Chr7	Marker8 Chr1	Marker9 Chr9	Marker10 Chr3
Ky21_1	147.564	GG	AA	CC	AA	AA	GG	AA	AA	GG	GG
Ky21_2	143.83	GG	AA	CC	GG	GG	GG	AA	AA	GG	GG
Ky21_3	148.83	GG	AA	CC	AA	CC	GG	AA	AA	GG	GG
Ky21_4	153.176	GG	AA	CC	AA	TT	GG	AA	AA	GG	GG
Oh43_1	149.677	AA	AA	CC	AA	GG	GG	AA	AA	AA	GG
Oh43_2	147.702	GG	AA	CC	AA	GG	GG	AA	AA	AA	GG
Oh43_3	143.899	GG	AA	CC	AA	GG	GG	AA	AA	AA	GG
Oh43_4	151.764	GG	AA	CC	AA	GG	GG	AA	AA	AA	GG
CML228_1	114.159	AA	GG	CC	GG	GG	TT	AA	AA	AA	CC
CML228_2	117.968	AA	GG	CC	GG	GG	CC	AA	AA	AA	AA
CML228_3	121.654	AA	GG	CC	GG	GG	AA	AA	AA	AA	GG
CML228_4	108.504	AA	GG	CC	GG	GG	GG	AA	AA	AA	AA
Hp301_1	121.261	AA	GG	CC	GG	GG	AA	CC	GG	AA	AA
Hp301_2	134.448	AA	GG	CC	AA	GG	GG	CC	GG	AA	AA
Hp301_3	135.971	GG	GG	CC	GG	GG	TT	CC	GG	AA	AA
Hp301_4	126.119	AA	GG	CC	GG	GG	CC	CC	GG	AA	AA

Take Marker3 as an example. All plants have the same SNP type CC at Marker3, but have different leaf angles. So, we don't expect that this marker is related to the leaf angle phenotype. Now, you can look for markers with SNP types that are exclusive for large or small leaf angles.

Next, you can conduct a *t-test* to determine if the leaf angles of individuals grouped by their SNP type at a given marker locus are significantly different.

A *t-test* is used to determine if there is a significant difference between the means of two groups. *t-test* is one of many tests used for the purpose of hypothesis testing in statistics.

Here, use Marker7 as an example. There are two SNP types, AA and CC, at Marker7.

Null Hypothesis ( $H_0$ ): The leaf angle mean of group AA is equal to the mean of group CC.

Alternative Hypothesis ( $H_a$ ): The leaf angle mean of group AA is not equal to the mean of group CC.

How to do the *t-test*...

*t-test* can be performed in Excel with the function *T.TEST()*.

- Input your data into Microsoft Excel.
- Type *=T.TEST()* in a cell.
- Select the measurements. Since you are comparing the leaf angle measurements between SNP types, select the leaf angle data from group AA for "array1", and then data points from group CC for "array2".
- Put 2 in "tails", since it is a two-side hypothesis test and 2 in "type".

	A	B	C	D
1	SeedlingID	2nd leaf Angle (°)	Marker7	
2			Chr7	
3	Ky21_1	147.564	AA	
4	Ky21_2	143.83	AA	
5	Ky21_3	148.83	AA	
6	Ky21_4	153.176	AA	
7	Oh43_1	149.677	AA	
8	Oh43_2	147.702	AA	
9	Oh43_3	143.899	AA	
10	Oh43_4	151.764	AA	
11	CML228_1	114.159	AA	
12	CML228_2	117.968	AA	
13	CML228_3	121.654	AA	
14	CML228_4	108.504	AA	
15	Hp301_1	121.261	CC	
16	Hp301_2	134.448	CC	
17	Hp301_3	135.971	CC	
18	Hp301_4	126.119	CC	
19			=T.TEST(B3:B14,B15:B18,2,2)	
20			T.TEST(array1, array2, tails, type)	

If your p-value is less than 0.05, this indicates there is less than 5% chance that the difference between the two groups is due to random chance. Then, you can reject our null hypothesis ( $H_0$ ) and conclude that there is a significant difference in leaf angle between SNP type AA and CC, which suggests that the two SNP types are correlated with the variation in leaf angles.

Q5. Is there a significant difference in leaf angles between the two SNP types at the Marker7 locus?

Look back at the SNP marker table. Based on the information provided, which SNP marker is 100% correlated with leaf angle? In other words, which marker has a SNP type exclusive to large angled plants and a different SNP type exclusive to small angled plants?

Q6. Performing the t-test again, do you see a significant difference for leaf angle between these two SNP types at this marker?