
Learning Objectives:

This assignment is motivated by a desire to better understand the nature of phenotypic variation associated with bananas, and the sources of said variation. From a learning objective perspective, this project will demonstrate your ability to i) manage data ii) perform a variety of different statistical tests iii) interpret the output of such tests iv) plot beautiful, meaningful figures/infographics v) present your findings to an interested audience. The required format for this assignment will be a narrated slide-deck, which is described in a later section of this document.

Project motivation / Research Questions

Bananas (*Musca spp.*) are a fruit consumed worldwide and the fourth most important food crop (after rice, wheat and maize) in the world (Prabha & Kumar 2015). Humans has been cultivating bananas for at least 8000 years (and possibly over 10,000 years)! Bananas are grown in over 150 countries each year with over 100 million metric tonnes of fruit being produced ear year. The most commonly internationally exported type of banana are those from the “Cavendish” cultivars belong to the AAA genome group, which includes the cultivars that have three sets of chromosome inherited from the wild (ancestral) species *Musa acuminata*. The bananas develop from flowers that grow on stems of this large, herbaceous (non-woody) plant. Bananas make up ~9% of the >6.5 billion dollars of fresh fruit imported in to Canada, (Agriculture and Agri-Food Canada, 2017), with ~ 15.70 kg of bananas consumed on average by each Canadian annually. Given their ubiquitous nature, bananas provide a great opportunity for us as biologists/statisticians to explore a number of different questions that are of great relevance from both biological and economic perspectives.

In this project, you use your own individualized banana dataset, and analyze it using the most appropriate statistical analysis techniques. The questions you will be trying to address are as follows:

- 1) How can we efficiently describe a banana’s phenotype? Does the phenotype of bananas depend on where (or) how they are bought?
- 2) Are there differences in the characteristics of groups of bananas that have come to Canada from different countries (or) by different companies? How much variation is there between/within these groups?
- 3) What is the relationship between a banana’s length and its width? Does it differ between those bananas on the inner row and the outer row?
- 4) Do the conditions under which bananas are stored affect how much it ripens over the course of a week?

How Will I Do this Assignment?

You should begin by visiting the course MyLS page to retrieve your personalized/individualized dataset (in csv format). The data therein contains a subset of the complete data collected by you and your classmates. The subset was created by me (using R of course) using a simple random

sampling technique to create a unique dataset for your personal use (i.e. everyone will have a slightly different set of data). To further individualize each of your assignments, the response and explanatory variables you will be analyzing will depend on when you were born & your student ID). Please consult the table below to determine which response variables you will be using.

Question 1: *How can we efficiently describe a banana's phenotype? Does the phenotype of bananas depend on where (or) how they were bought?*

1. Using your individualized dataset, perform a Principle Component Analysis using the following variables: i) Outer curve length ii) Inner curve length, iii) Pedicel length, iv) Banana Circumference, v) Pedicel circumference, and vi) Number of sides.
2. Report in a table the loadings and % variance explained by the first 3 principle components. Depending on your birth month (see below), describe ONE of the principle components via your interpretation of its loadings
3. Using your re-scored Principle Component scores from your chosen component, as the response variables, determine statistically whether or there is a difference in the mean (or) median scores for bananas of the bananas bought in Kitchener-Waterloo or outside of K-W. Remember to first determine whether or not, for each of the two samples, whether the variables being compared have a normal distributions. Transform, in an appropriate fashion when the data is not be normally distributed, and determine whether the transformed data is/is not normally distributed. Also remember to first test whether the variances in the two samples are equal and choose the most appropriate parametric or non-parametric statistical test based on this information to make the comparison of means (or) medians.
4. Calculate the (appropriate) effect size statistic (& 95% CI) of this difference
5. Perform a two-sample power test to determine what the current power of your one-sample test is as well as determine the minimum sample size for a one sample test that would be necessary to detect a difference of this magnitude (with power = 0.8).
6. Using the graphing functions of R, create a simple box plot graph that summarizes the comparison being made in step 4 above.

Question 1	Variable used to describe banana "size"
Birthday in January, February, March or April	Principle Component 1
Birthday in May, June, July or August	Principle Component 2
Birthday in September, October, November or December	Principle Component 3

Question 2: *Are there differences in the characteristics of groups of bananas that have come to Canada from different countries (or) by different companies? How much variation is there between/within these groups?*

1. Consult the tables below to determine your specific response and explanatory variables.
2. Using a one-way fixed-effect ANOVA (or Kruskal-Wallis test) compare the mean/medians in your chosen variable measured by different groups. If test is significant, use post-hoc means comparison methods to determine exactly which groups differ from which.
3. Produce a boxplot figure illustrating the data for by each group (don't forget to properly indicate, on your graph, if necessary, which means/medians (if any) are significantly different using connected letters.
4. Using a one-way random-effects ANOVA, determine how much of the variance in your chosen variable can be attributed to between group variation, and how much can be attributed to within group variation.

Question 2 (Part 2)	Response variable that describes the banana's "characteristic"
Born on the 1 st through 6 th of the month	Pedicle Length
Born on the 7 th through 12 th of the month	Pedicle Circumference
Born on the 13 th through 18 th of the month	USDA Ripeness Index Value (Start)
Born on the 19 th through 25 th of the month	Wright Ripeness Index Value (Start)
Born on the 26 th through 31 st of the month	Banana Circumference

Question 2 (Part 2)	"Explanatory" variable that describes the banana's origin/importer
Last digit in your student ID number is odd,	Country of Banana Origin
Last digit in your student ID number is even	Banana Import Company

Question 3: *What is the relationship between a banana's length and its width?*

1. Consult the tables below to determine your specific X and Y variables.
2. Calculate the appropriate correlation coefficient statistic for the relationship between your two variables, as well as determine if it is significantly different from zero.
3. Be sure to include any information relevant to your analysis (i.e. tests of normality, homogeneity of variance etc..., and/or any relevant transformation(s) performed) that influenced your choice of statistical test.
4. Produce a scatterplot + regression line displaying the relationship between your two variables.
5. Calculate the slope and the intercept of your regression line, and determine if these two parameters are significantly different from zero.

Question 3 (<i>X variable - width</i>)	“X” Variable
Birthday in January, February, March or April	Banana Diameter
Birthday in May, June, July or August	Pulp Diameter
Birthday in September, October, November or December	Banana Circumference

Question 3 (<i>Y variable - length</i>)	“Y” Variable
Next-to-last digit in your student ID number is odd	Outer Curve Length
Next-to-last digit in your student ID number is even	Inner Curve Length

Question 4: *Do the conditions under which bananas are stored affect how much it ripens over the course of a week?*

1. Perform a two-way ANOVA to determine if light treatment and/or spacing treatment have a significant effect on the in ripeness index (End index value – start index value).
2. Report & interpret the results of your ANOVA table.
3. Create a plot that best depicts in your opinion the most important results of your experiment, indicating the location of any statistically significant differences revealed by your statistical tests.
- 4.

Question 4	Response Variable
Your birth year is an odd number	USDA Difference
Your birth year is an even number	Wright Difference

Report Format (Video!)

For this assignment, you will be submitting your report (electronically) in a short/small narrated powerpoint video+ a written abstract that summarizes your study + a document that lists your annotated R codes. The slides should present relevant information tables/figures/(relevant)statistical results and a 1-2 line summary of what is being presented (see specific details outlined below), while your narration will explain what is being presented on each slide. The whole presentation should be no more than 9 slides (*see below*) and ~10-12 minutes.

Here is a rough breakdown of what each of these slides should contain. Your voiceover should guide the viewer through your statistical decision-making process, your results and their interpretation.

Slide 1. Title, your name and a picture of a banana ☺

*Slide 2. **Question 1 (part 1):** How can we efficiently describe a banana's phenotype?*

Present a table with the % variance explained, and the loadings table for the first 3 principle components that you obtained when conducting your PCA. For your specific PC, please describe/interpret the loadings (ie. What does a large, positive PC score represent vs. a large, negative PC score?).

*Slide 3. **Question 1 (part 2):** Does the phenotype of bananas depend on where (or) how they were bought?*

Display side-by-side boxplots of the re-scored PC values for the bananas bought in Kitchener-Waterloo & those bought outside K-W.

Report the results of a two-sample test that compares the means (or) medians of these groups. Include results/interpretation the tests the data's normality/equality of variance that influenced your choice of two-sample test.

Report the (appropriate) effect size statistic (& 95% CI) of this difference, as well as a two-sample power test to determine what the current power of your one-sample test is as well as determine the minimum sample size for a two sample test that would be necessary to detect a difference of this magnitude (with power = 0.8).

*Slide 4. **Question 2 (part 1):** Are there differences in the characteristics of groups of bananas that have come to Canada from different countries (or) by different companies?*

Present a boxplot figure illustrating the distribution of data for by each group. Report the results of a one-way fixed-effect ANOVA (or Kruskal-Wallis test) that compares the mean/medians in your chosen variable measured by different groups. If the test was significant, use post-hoc means comparison methods to determine exactly which groups differ from which, and display the location of these differences on the plot using the connected letters method

Be sure to include relevant information on normality/equality of variance that was relevant to your decision what group comparison test was most appropriate for your data.

*Slide 5. **Question 2 (part 2):** How much variation is there in the characteristics of groups of bananas between/within these groups?*

Present a table illustrating the output of one-way ANOVA using your variables. Explain (using the random-effects method) how much of the variation can be attributed to within-group versus between-group effects.

*Slide 6. **Question 3:** What is the relationship between a banana's length and its width?*

Present a scatterplot using your specific X & Y variables. Report the appropriate correlation coefficient (and its statistical significance).

Be sure to include relevant information on normality/equality of variance that was relevant to your decision what group comparison test was most appropriate for your data.

Report the slope and intercept of a regression line through the data, as well as the statistical significance of those parameters. Plot the regression line on the scatterplot.

Slide 7. **Question 4:** *Do the conditions under which bananas are stored affect how much it ripens over the course of a week?*

Report the results of a two-way ANOVA to determine if light treatment and/or spacing treatment have a significant effect on the in ripeness index (End index value – start index value), and interpret the results of your ANOVA table.

Slide 8. **Question 4 (continued):** *Do the conditions under which bananas are stored affect how much it ripens over the course of a week?*

Present a plot that best depicts in your opinion the most important results of your experiment, indicating the location of any statistically significant differences revealed by your statistical tests.

Slide 9. Acknowledgements. Did you work with anyone, are there any resources you would like to reference, or anyone you want to thank?

Annotated example of an abstract taken from Nature 435, 114-118 (5 May 2005)

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline.

Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines.

One sentence clearly stating the **general problem** being addressed by this particular

study.

One sentence summarising the main result (with the words "**here we show**" or their equivalent).

Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge.

One or two sentences to put the results into a **more general context**.

Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline, may be included in the first paragraph if the editor considers that the accessibility of the paper is significantly enhanced by their inclusion. Under these circumstances, the length of the paragraph can be up to 300 words. (The above example is 190 words without the final section, and 250 words with it).

During cell division, mitotic spindles are assembled by microtubule-based motor proteins^{1,2}. The bipolar organization of spindles is essential for proper segregation of chromosomes, and requires plus-end-directed homotetrameric motor proteins of the widely conserved kinesin-5 (BimC) family³. Hypotheses for bipolar spindle formation include the 'push-pull mitotic muscle' model, in which kinesin-5 and opposing motor proteins act between overlapping microtubules^{4,5}. However, the precise roles of kinesin-5 during this process are unknown. Here we show that the vertebrate kinesin-5 Eg5 drives the sliding of microtubules depending on their relative orientation. We found in controlled *in vitro* assays that Eg5 has the remarkable capability of simultaneously moving at $\sim 20 \text{ nm s}^{-1}$ towards the plus-ends of each of the two microtubules it crosslinks. For anti-parallel microtubules, this results in relative sliding at $\sim 40 \text{ nm s}^{-1}$, comparable to spindle pole separation rates *in vivo*⁶. Furthermore, we found that Eg5 can tether microtubule plus-ends, suggesting an additional microtubule-binding mode for Eg5. Our results demonstrate how members of the kinesin-5 family are likely to function in mitosis, pushing apart interpolar microtubules as well as recruiting microtubules into bundles that are subsequently polarized by relative sliding. We anticipate our assay to be a starting point for more sophisticated *in vitro* models of mitotic spindles. For example, the individual and combined action of multiple mitotic motors could be tested, including minus-end-directed motors opposing Eg5 motility. Furthermore, Eg5 inhibition is a major target of anti-cancer drug development, and a well-defined and quantitative assay for motor function will be relevant for such developments.

Some technical advice on how to make a narrated video presentation:

<https://www.youtube.com/watch?v=Y5dgwwa5XRA>

<https://www.youtube.com/watch?v=D8JV3w4TOVw>

<https://www.youtube.com/watch?v=qcaHMbKGnRY>

<https://www.ispringsolutions.com/blog/how-to-create-a-video-presentation-with-ispring-presenter>

https://support.microsoft.com/en-us/office/record-a-presentation-2570dff5-f81c-40bc-b404-e04e95ffab33?wt.mc_id=aid573689_qsg_141923&ui=en-us&rs=en-us&ad=us

<https://support.microsoft.com/en-us/office/record-a-slide-show-with-narration-and-slide-timings-0b9502c6-5f6c-40ae-b1e7-e47d8741161c#officeversion=2013, 2016>

<https://support.microsoft.com/en-us/office/turn-your-presentation-into-a-video-c140551f-cb37-4818-b5d4-3e30815c3e83>

Please remember that you have free access to Microsoft 365: <https://students.wlu.ca/services-and-spaces/tech-services/software/microsoft-office-365.html>

If you want to work in Keynote: <https://support.apple.com/en-ca/guide/keynote/tan8a5df9cc5/mac>

<https://smallbusiness.chron.com/recording-slideshow-keynote-33813.html>

<https://www.youtube.com/watch?v=C6e2ZtHnimA>

Rubrics are matrixes that define what is expected in a learning situation. They offer an opportunity to evaluate the student's understanding of a scientific topic by levels of performance on certain criteria. They can make assessment more meaningful, clarify expectations, and yield better feedback. Your assignment will be graded out of 30 points using the following rubric.

Please note: Reports lacking in organization, multiple spelling/grammar errors will be docked up to 3 points.

	Points
Abstract 4 pts.	A one-paragraph summary of your study's objectives and key findings (see above for a great example of an abstract).
Question 1 <i>How can we efficiently describe a banana's phenotype?</i> 3 pts.	Principle component table (with loadings & % explained) present. Clear and accurate interpretation of your specific principle component's loadings.
Question 1 <i>Does the phenotype of bananas depend on where (or) how they were bought?</i> 5 pts.	The figure clearly depicts the data from your study in an aesthetically pleasing visual manner. Axes are clear & meaningful The results of the appropriate two-sample statistical tests are included (and correctly reported/interpreted). Results of relevant tests that influenced the decision to perform that test are also presented. Appropriate effect size statistic (& 95% CI) is presented. Results of two-sample power test to determine what the current power of your one-sample test is as well as determine the minimum sample size for a two-sample test that would be necessary to detect a difference of this magnitude (with power = 0.8). All results are clearly presented and accurately interpreted.
Question 2 <i>Are there differences in the characteristics of groups of bananas that have come to Canada from different countries (or) by different companies?</i> 3 pts.	The figure clearly depicts the data from your study in an aesthetically pleasing visual manner. Axes are clear & meaningful. Connected letters used to indicate location of any significant differences in means (or) medians. The results of the appropriate 3+ group statistical comparison of group means (or) medians is presented. Results of relevant tests that influenced the decision to perform that test are also presented. All results are clearly presented and accurately interpreted.

<p>Question 2 <i>How much variation is there between/within these groups?</i></p> <p>3 pts.</p>	<p>Present one-way ANOVA table generated using your variables.</p> <p>Explain (using the random-effects method) how much of the variation can be attributed to within-group versus between-group effects.</p> <p>All results are clearly presented and accurately interpreted.</p>
<p>Question 3: <i>What is the relationship between a banana's length and its width?</i></p> <p>5 pts.</p>	<p>Scatterplot of + regression line using your specific X & Y variables. Axes are clear & meaningful.</p> <p>Report the appropriate correlation coefficient (and its statistical significance). Results of relevant tests that influenced the decision to perform that test are also presented.</p> <p>Report the slope and intercept of a regression line through the data, as well as the statistical significance of those parameters.</p> <p>All results are clearly presented and accurately interpreted.</p>
<p>Question 4: <i>Do the conditions under which bananas are stored affect how much it ripens over the course of a week?</i></p> <p>5 pts.</p>	<p>Present a two-way ANOVA table depicting analysis to determine if light treatment and/or spacing treatment have a significant effect on the in ripeness index.</p> <p>Present a plot that best depicts in your opinion the most important results of your experiment, indicating the location of any statistically significant differences revealed by your statistical tests.</p> <p>All results are clearly presented and accurately interpreted.</p>
<p>Annotated R codes</p> <p>2pts</p>	<p>Annotations (descriptions of what code done). . Just one example of each function. Do not include R output - just functions + annotation.</p>