Implementing Specifications Grading Into College-Level Courses That Use Projects

Jason Elsinger, Ph.D.

Department of Mathematics
Lewis University

SIMIODE EXPO
February 9, 2024

Criteria for Evaluating Grading Systems

1. Uphold high academic standards

$$\downarrow \text{rigor} \quad \Longrightarrow \quad \uparrow \text{student satisfaction}$$

2. Reflect the learning outcomes

Which parts of the course?

Which learning outcomes exceptionally well or not at all?

3. Motivate students to learn (vs. performance orientation)

points-based system $\Longrightarrow$ turns education into a game

4. Motivate students to excel

built-in incentives to demand strong performances to earn credit

5. Reduce stress
6. Make students feel responsible for their learning (earned vs. given)

References on Specs and Standards-Based Grading

Grading for Growth by David Clark and Robert Talbert

Specifications Grading by Linda B. Nilson

How Learning Works by Ambrose, Bridges, Lovett, DiPietro, Norman

Learner-Centered Teaching: Five Changes to Practice by Maryellen Weimer

Applying a Standards-Based Grading Framework Across Lower Level Mathematics Courses (with D. Lewis)
`https://www.tandfonline.com/doi/full/10.1080/10511970.2019.1674430`

Specifications grading into college mathematics courses using a points-based system (In Preparation)

Main Features of Specs Grading

1. Students are provided with a clear list of learning objectives (LOs)
2. Final grades are based on how many LOs are successfully completed
3. Students are provided ample opportunities to reassess previous unsuccessful LOs
4. Attempts are graded using a 0-3 scale, and only the best mark is kept
5. Students can master a LO up to two times, yielding a max of 6 knowledge points

Institutions and Courses

Courses:

- Contemporary Math
- Differential Equations
- Calculus 3
- Cryptology and its History
- Precalculus
- Geometry and its History
- Linear Algebra
- Business Calculus
- Elementary Statistics
- Calculus 1
- College Algebra
- Probability and Statistics

Florida Southern College:

- Private, Liberal arts
- Around 2,400 students
- Class sizes: 13, 25

Course Design by Learning Outcomes (Statistics)

Course is split into separate standards grouped into 4 Big Questions.

- Sampling: Where does data come from?
- Organization: How can data be organized best?
- Probability: How do we describe chance?
- Inference: How do we draw conclusions from data?

Each module contains individual learning outcomes.

Sampling: Where does data come from?

1.1. **Definitions.** Describe an example of a statistic or parameter, or answer a conceptual question.

1.2. **Descriptive Literacy.** Describe the population, sample, variable, statistic, or parameter for a given situation.

1.3. **Sampling Techniques.** Indicate the sampling technique used in a given experiment or give an example that demonstrates a technique. Describe the steps to draw a simple random sample, and use random numbers to showcase the steps.

1.4. **Bias.** Determine if a sample collected contains bias and indicate clearly what that bias is.

Determining Final Grades (Elementary Statistics)

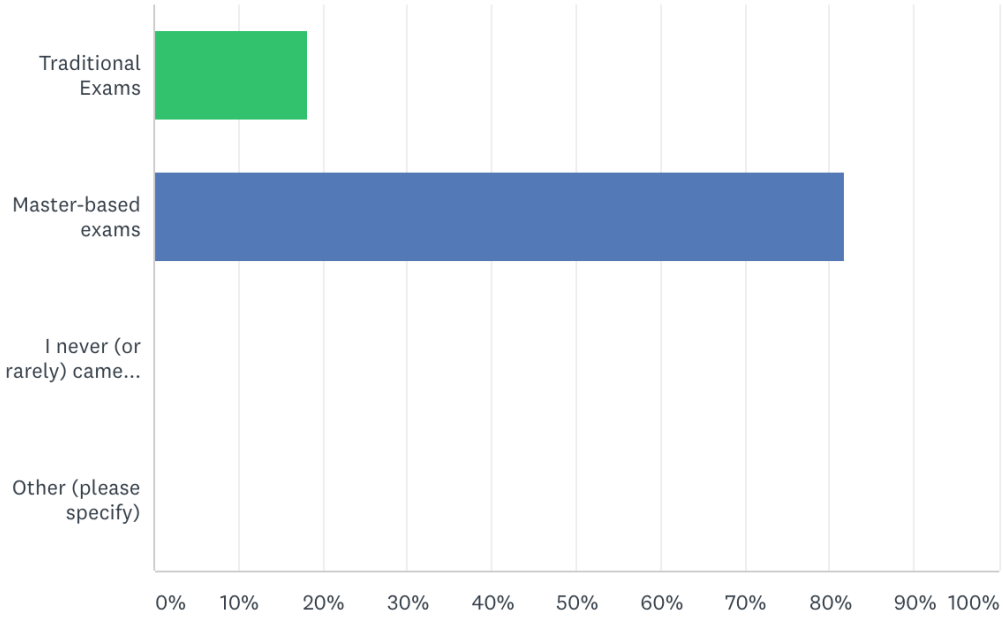
Homework points vs. Knowledge points

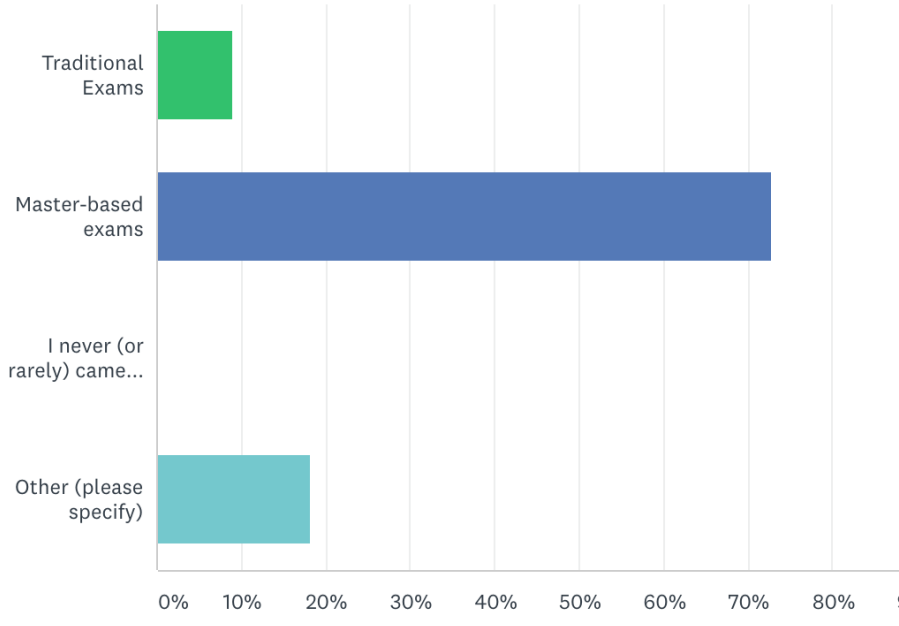
To earn a...	...earn at least the following point totals
A	135 knowledge points from individual problems, and 82 homework points from homework and mini-projects
B	111 knowledge points from individual problems, and 62 homework points from homework and mini-projects

(27 LOs = 162 possible; 90% = 146, 80%=130)

Advantages

- gives students more autonomy
- students learn from their mistakes
- students became more aware of simple mistakes
- writing improved
- students attended office hours and took them seriously
- effects of partial credit reduced





Oreo Project (Elementary Statistics)

Do Double Stuf Oreos really have twice the filling?

- Collected data in-class using knives and scales
- Google sheets for real-time data collection

Google Doc (can use for groups)
`https://docs.google.com/document/d/1ky781aBvevqWeMS_vZW8UY1y8rMOr9rX8O_xLWnHexg/edit`

Use a Google Sheet to collect data from everyone in real time!

Making Data Meaningful (Parts 1 & 2)
`https://unece.org/statistics/publications/making-data-meaningful-part-1`

Criteria for an A

Overall (Final Submission)

- Title page, page numbering, catchy but accurate title
- Abstract
- Proper grammar throughout
- Appropriate length in every section
- All figures/graphs properly labeled
- References (more than one)
- In-line citations

Example of A Work

Running Head: ARE DOUBLE STUF REALLY DOUBLE THE STUF

Are Double Stuf Really Double the Stuf?

Example of A Work

Abstract

Our class set out to test if Double Stuf Oreos really contained double the filling compared to the traditional Oreos. Oreos were invented in 1912 and after 62 years, they launched their Double Stuf variety for those who love the filling. Since then, Nabisco has held onto the secret of exactly how much filling is in each cookie. Our class weighed out the filling of 52 traditional Oreos and 40 Double Stuf to find out. We set the claim that they would contain exactly twice the filling and used a level of significance of 5%. After averaging the weights and performing a two sample t test, we found the cookies contained on average 2.1 times the filling and our p value of >0.001 confirmed that this statistic was significant enough to rule out the Double Stuf cookies as containing exactly twice the filling.

Criteria for a B

Overall (Final Submission)

- Title page, page numbering, catchy title
- Abstract
- Proper grammar throughout
- Appropriate length in every section
- All figures/graphs properly labeled
- Reference (more than one)
- In-line citations

Criteria for a C

Overall (Final Submission)

- Title page, page numbering, some title
- Abstract
- Decent grammar throughout
- Appropriate length in every section
- All figures/graphs properly labeled
- References (more than one)
- In-line citations

Criteria for an A

Introduction (Part 1)

- History of Oreos, and the introduction of different types
- Reference/discuss previous studies
- Description of the question
- Describes purpose of the study
- Appropriate length (1-2 pg, double spaced)

Example of A Work

Introduction

Originally in 1908, the Sunshine company released its "Hydrox" cookie. The first sandwich cookie of its kind. Later in 1912, the Nabisco company launched its imitation cookie: The "Oreo Biscuit", a cookie with a sweeter filling and a less cruncher cookie shell. The new product quickly exceeded the popularity of the original Hydrox cookie, even being names the "Best selling cookie of the 20th century," and has led to the Hydrox cookie being thought of as the imitator. In 1921 and 1948, the cookie underwent two name changes, first to "Oreo sandwich" then to "Oreo Cream Sandwich" and finally in 1974, the company landed on "Oreo Chocolate Sandwich Cookie." The first flavored Oreo to enter the market was a lemon flavor introduced in the 1920's only to later be discontinued. In 1925 Nabisco, launched the "Oreo Double Stuf Sandwich Cookies." These cookies, by their name, claim to contain twice the "stuf" or filling of the original Oreo cookies. This claim has been put to the test before in 2013 by a high school class led by teacher Dan Anderson and again by Butterworth et al. (2014) and is being put to the test again by our class.

According to Mr. Anderson's class results, they determined there was in fact not double the "stuf" in the Double Stuf Oreos compared to traditional cookies. The class came to this

Criteria for a B

Introduction (Part 1)

- History of Oreos, and the introduction of different types
- Reference/discuss previous studies
- Description of the question
- Describes purpose of the study
- Appropriate length (1/2 - 1 pg, double spaced)

Example of B Work

Introduction

The purpose and claim of this study is to look at Oreos and see if the Double-Stuffed Oreos have double the normal Oreo amount of stuffing/icing. This is done by weighing the icing of normal Oreos and comparing a sample of the normal Oreos to the sampled of the weights of the stuffing of Double Stuffed Oreos. The weight (grams) of the stuffing was found by scarping the insides of two normal Oreos separately and weighing them in a paper cup with negligible weight (0.1 grams), and then repeating this with the Double-Stuffed Oreos. The sample was more random as cookies were handed out indiscriminately by Professor Elsinger, but to make the sample more random, one would need to get Oreos from each country they are available, and from different stores, as the cookies might be made differently and the weight may have changed in how they were handled. This would be very difficult to do.

Criteria for a C

Introduction (Part 1)

- History of Oreos, and the introduction of different types
- Reference/discuss previous studies
- Description of the question
- Describes purpose of the study
- Appropriate length (1 paragraph, double spaced)

Example of C Work

Introduction

The purpose of this study is to test the claim that Double Stuf Oreos contain twice as much filling as Traditional Oreos. We are interpreting "twice the filling" by weight. The unit used was grams and the filling was obtained by scraping it off Oreo cookies. To get a simple random sample, one could, without looking, take the wanted number of cookies out of the box. It is feasible because it contains no bias.

Example of Not Yet Passing

Oreo project

Introduction

In this study as a class, we weighed the filling of Oreos in order to test the claim that Double Stuffed Oreos contain twice as much filling as Traditional Oreos.

Criteria for an A

Results (Part 2)

- Includes plots (boxplots, histograms, QQ-plots)
- Describes the shape and other relevant features of plots
- Presents quantitative statistics clearly
- Describes/Interprets the quantitative statistics
- Assesses the assumption of normality
- Show details of the hypothesis test (added later)
- Conclude the hypothesis test correctly (added later)

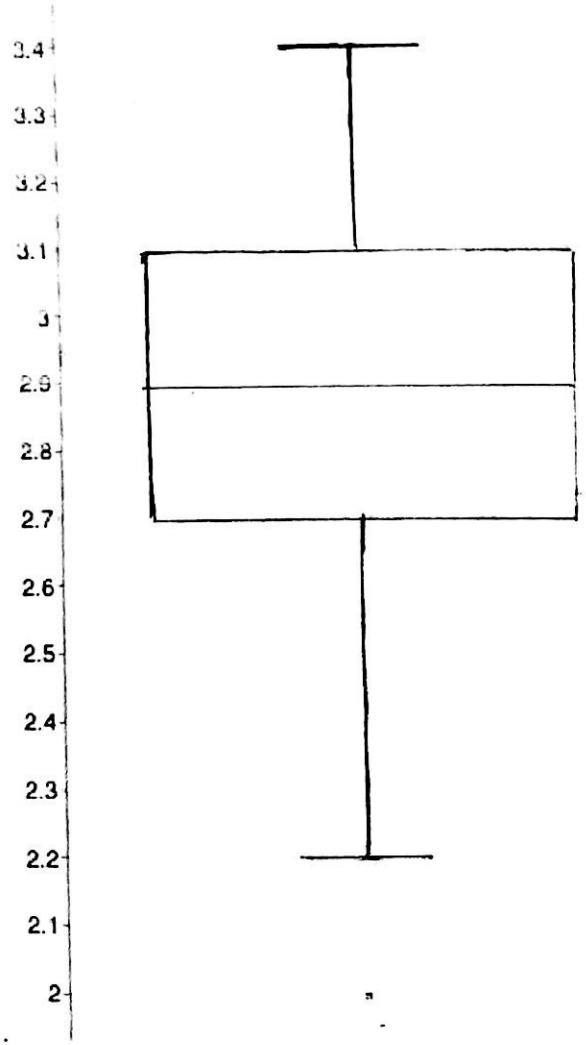
Criteria for a B

Results (Part 2)

- Includes plots (boxplots, histograms, QQ-plots)
- Describes the shape and other relevant features of plots
- Presents quantitative statistics clearly
- Describes/Interprets the quantitative statistics
- Assesses the assumption of normality
- Show details of the hypothesis test (added later)
- Conclude the hypothesis test correctly (added later)

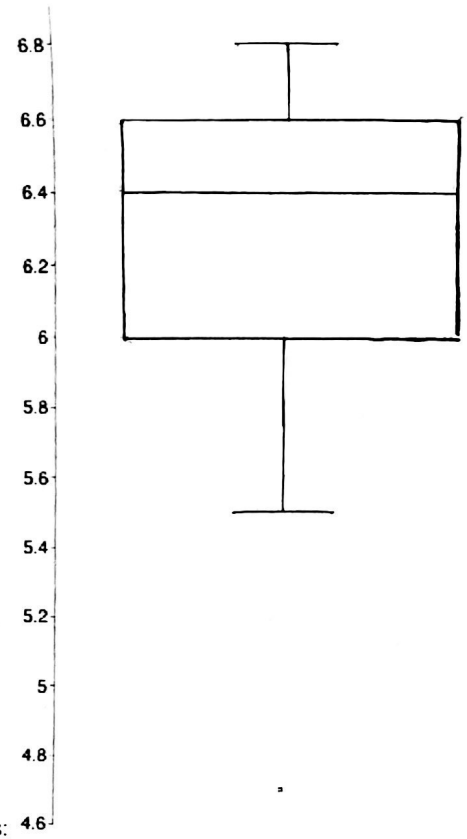
Example of B Work

Results



Traditional Oreos:

Example of B Work



Double Stuf Oreos:

	Number of cookies	Average	Median	Standard Deviation	Minimum	Maximum
Traditional	48	2.9	2.9	0.3	3	3.4
Double Stuf	46	6.3	6.4	0.4	4.7	6.8

Example of B Work

The P-value is < 0.05 so we reject the null hypothesis. In conclusion there is not sufficient evidence to support that Double Stuf Oreos contain twice as much filling as Traditional Oreos. The box and whisker graphs both show that there is an outlier which affects the mean of the data. The Traditional Oreo box and whisker plot shows the data has a symmetric distribution because the median and the average equal each other which are both 2.9 which can be found in the table below the graphs. The Double Stuf Oreos data is more skewed left and with this you can tell by looking at the box and whisker plot because most of the data is on the right side, but also by looking in the table because the median is bigger than the average which indicates that the distribution is skewed left.

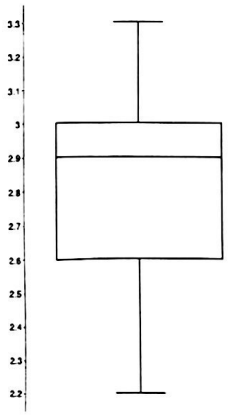
Criteria for a C

Results (Part 2)

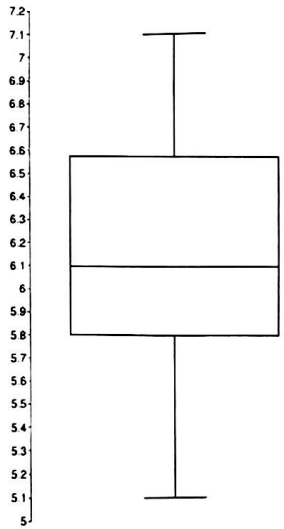
- Includes plots (boxplots, histograms, QQ-plots)
- Describes the shape and other relevant features of plots
- Presents quantitative statistics clearly
- Describes/Interprets the quantitative statistics
- Assesses the assumption of normality
- Show details of the hypothesis test (added later)
- Conclude the hypothesis test correctly (added later)

Example of C Work

Results



This is the box and whisker plot for the sample of Traditional Oreos.



This is the box and whisker plot for the sample of Double-Stuffed Oreos.

Example of C Work

To see calculations, data, and the t-test, refer to the attached Oreo report (Both Parts 1 and 2).

For the Traditional Oreos, the graph shows that data is skewed right, and for the Double-Stuffed Oreos, the graph shows that the data is skewed left. The difference in standard deviations shows that Traditional Oreos sample has less spread than the Double Stuffed Oreos.

Example of Not Yet Passing

Results

After everything was calculated, the null hypothesis was rejected which meant that there was no support for the claim that Double Stuffed Oreos contain twice as much filling as Traditional Oreos.

Criteria for an A

Methodology (Part 3)

- Describes sampling procedure, units, etc.
- Describes the statistical test to be used (and P-value or CI method)
- Describes the assumptions needed and necessary formulas
- Revisit QQ-plots and assessment of normality
- Describes hypotheses, level, and power used clearly
- Describes the possible errors that occur (type I and II)

Criteria for a B

Methodology (Part 3)

- Describes sampling procedure, units, etc.
- Describes the statistical test to be used (and P-value or CI method)
- Describes the assumptions needed and necessary formulas
- Revisit QQ-plots and assessment of normality
- Describes hypotheses, level, and power used clearly
- Describes the possible errors (types I and II)

Example of B Work

Methodology

The two populations we studied were the fillings inside Traditional Oreos and Double Stuffed Oreos. The sample of cookies was obtained by purchasing them from Publix. In the study, we took the samples of Double Stuffed Oreos and Traditional Oreos out of the packaging then gave everyone in the class at least two Oreos from each sample. As a class, we then proceeded to scrape the fillings out of each Oreo with a plastic knife into separate small plastic containers and weighed them on a small scale. We also weighed each filling, Traditional and Double Stuffed, separately by first giving everyone the Traditional Oreos, scraped out the filling of each Traditional Oreo we were given, put each one into a different small plastic container weighing each filling individually, and then after weighing at least two Traditional Oreo fillings we then proceeded to do the exact same steps with the Double Stuffed Oreos using different individual small containers but the same plastic knife to scrape the filling out. After collecting all of the data, as a class, we put all of the weights of the Oreo fillings in excel. In the excel sheet, we put all the weights for the Oreos like this:

Example of B Work

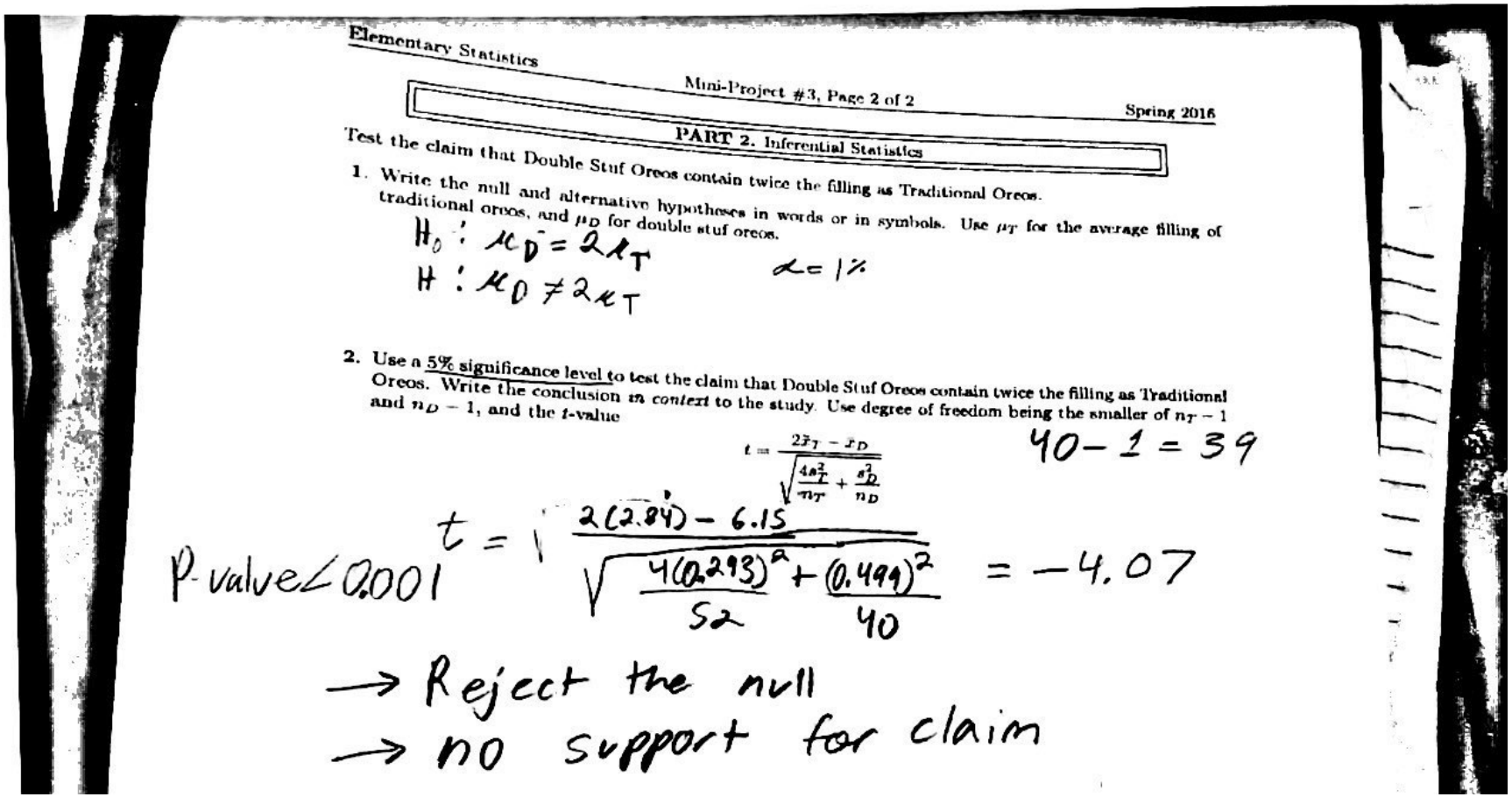
	A	B	C	D	E	F	G	H	I	J
2	Student	Oreo #1	Oreo #2	Oreo #3	Oreo #4		Oreo #1	Oreo #2	Oreo #3	Oreo #4
3	Kalli	3	2.5	2.4	3		5.7	6	7.1	
4	Bethany	2.7	2.6	2.9	3.3		5.4	6.1		
5	Calen	2.7	2.5	2.7			6.2	5.8	6.4	
6	Tea	2.9	3.3	3.1			6.2	5.9		
7	Saroya	2.9	2.5							
8	Stefanie	3.3	3.2				5.8	5.9		
9	Ashley									
10	Ava	3.1	2.7	3.1			5.8	6.2		
11	Patrick	2.9	2.7				6.8	6.9		
12	Nadine									
13	Madisor	3	2.5				6.3	5.8		
14	Michelle									
15	Brett	2.9	2.9	2.9	3.1		6.6	5.7	5.8	6
16	AJ									
17	Spencer	2.6	3.2	3.2	2.3		6.8	6.8	6.7	
18	Joshua	2.5	2.3				5.1	5.5		
19	Dylan	2.7	2.2				5.1	6.1		
20	Alegdah	2.9	3				6.4	6.5		
21	Logan	2.6	3.1	3.1	3.1		6.3	6.4	6.9	
22	Jasmine	2.9	3.2							
23	Brandor	2.3	3.2				5.8	5.9		
24	Charlie	3	2.9	3			6.6	6.9		
25	Jacobus	2.7	2.6				6.1	5.7		

then proceed to count the number of cookies for Traditional and Double Stuffed Oreos,

calculated the mean, median, standard deviation, minimum, and maximum all using excel.

Example of B Work

The class wrote an alternative and null hypothesis for the claim. Using the numbers we got in excel and plugged them into the equation as shown below.



Elementary Statistics — Mini-Project #3, Page 2 of 2 — Spring 2016

PART 2. Inferential Statistics

Test the claim that Double Stuf Oreos contain twice the filling as Traditional Oreos.

1. Write the null and alternative hypotheses in words or in symbols. Use μ_T for the average filling of traditional oreos, and μ_D for double stuf oreos.

$H_0: \mu_D = 2\mu_T$

$H: \mu_D \neq 2\mu_T$ $\quad \alpha = 1\%$

2. Use a 5% significance level to test the claim that Double Stuf Oreos contain twice the filling as Traditional Oreos. Write the conclusion *in context* to the study. Use degree of freedom being the smaller of $n_T - 1$ and $n_D - 1$, and the t-value

$$t = \frac{2\bar{x}_T - \bar{x}_D}{\sqrt{\frac{4s_T^2}{n_T} + \frac{s_D^2}{n_D}}}$$

$40 - 1 = 39$

$$t = \frac{2(2.84) - 6.15}{\sqrt{\frac{4(0.293)^2}{52} + \frac{(0.499)^2}{40}}} = -4.07$$

P-value < 0.001

→ Reject the null

→ no support for claim

Criteria for a C

Methodology (Part 3)

- Describes sampling procedure, units, etc.
- Describes the statistical test to be used (and P-value or CI method)
- Describes the assumptions needed and necessary formulas
- Revisit QQ-plots and assessment of normality
- Describes hypotheses, level, and power used clearly
- Describes the possible errors that occur (type I and II)

Example of C Work

Methodology

The statistical test that was used in this experiment was the hypothesis test for average. The test was a two-tailed test because we were testing whether or not the average Double Stuf Oreo filling was twice as much as an average Traditional Oreo Filling. We used the symbol μ (Mu) which is a Greek letter meaning average. The subscripts of D and T stand for Double Stuf Oreo filling and Traditional Oreo filling respectively. Our hypotheses were as follows; H_0: $\mu_D = 2\mu_T$ and H_1: $\mu_D > 2\mu_T$.

Example of Not Yet Passing

Methodology

Using SPSS, and a two tailed test, the professor chooses the level of sig

H0: MD= 2MT (claim)

H1: MD=/ 2MT

The level of significance is 5%

Criteria for an A

Discussion (Part 4)

- Describes possible errors
- Interprets the results section properly
- Describes possible lurking variables
- Describes possible sources of bias

Recommendations (Part 4)

- Describes what could have been done differently
- Addresses potential issues

Criteria for a B

Discussion (Part 4)

- Describes possible errors
- Interprets the results section properly
- Describes possible lurking variables
- Describes possible sources of bias

Recommendations (Part 4)

- Describes what could have been done differently
- Addresses potential issues

Example of B Work

Discussion:

In terms of potential issues with these data sets, one glaring one is that the sample sizes are very small. This can cause a lack of reliability for the data as the Law of large numbers suggests that the larger the sample size, the closer to the expected value the average will become. Another issue is the large potential for bias within this experiment. The way the cookies were

obtained could be biased. A procedure that could've been done better was the scraping of the filling off of the cookie. In terms of the validity of these conclusions, one should always be suspicious of what they are being told to believe.

Recommendations:

If this experiment was to be done again, I would suggest that the cookies are obtained via a simple random sampling procedure to promote total randomization and an unbiased base set of data. I would also implement something to completely scrape all of the filling off of the cookies in a consistent manner, as there was residue on the cookie leftover from scraping.

Criteria for a C

Discussion (Part 4)

- Describes possible errors
- Interprets the results section properly
- Describes possible lurking variables
- Describes possible sources of bias

Recommendations (Part 4)

- Describes what could have been done differently
- Addresses potential issues

Example of C Work

Discussion

One error that was involved with this experiment was when the samples were collected. When the stuffing was scraped off the cookies, it was not always possible to get all the stuffing off, or in some cases some of the cookie was scraped off as well.
Another error that could occur was when weighting the samples. If, for instance, the one weighting the stuffing forgot to reset the scale with the cupcake-holders on them, the weight would be wrong.

Another error that could be involved was the knife. In some instances, the stuffing got stuck on the knife, and it was difficult to get it off. That means that in some samples there might have been some stuffing missing.
We should not be suspicious of our conclusion, because the package no longer says "twice the filling" which means that the company themselves say that every cookie will not have twice the stuffing. There might still be some cookies in every box that have twice the stuffing, but there is no certainty of it.

Recommendations

Things that can be done different is to make sure to always reset the scale and make sure that it is right, and to use a tool that the stuffing will not get stuck to when scraping.

Thank you!

Jason Elsinger jelsinger@lewisu.edu

Applying a Standards-Based Grading Framework Across Lower Level Mathematics Courses (with D. Lewis)
Elsinger, J., Lewis, D. (2020). PRIMUS, 30(8-10), 885-907.

Specifications grading into college mathematics courses using a points-based system (In Preparation)

Using technology to aid in implementing specifications grading (In Preparation)