A Hands-on Inferential Statistics Experiment: Do Students Perceive Value?

Sudipta Roy
University of St. Francis, 500 Wilcox Street, Joliet, IL 60435

Abstract
Published research in statistics education contain many examples of hands-on experiments that instructors use in their statistics classes. Clearly, instructors believe these experiments enhance student learning. But, do students perceive value? In this paper, students' perception of a hands-on inferential statistics experiment is assessed using a perception survey that addresses different dimensions of learning and engagement. The results indicate that students do find value in this in-class experiment. The study adds to the rich literature on statistical experiments in the classroom by providing a students' perspective on their use.

Keywords: Students' perceptions, in-class experiments, perception survey

1. Introduction

Journals in statistical education are rife with documentation of various kinds of experiments that teachers conduct in their classrooms. Pfaff & Weinberg (2009) designed experiments using cards and dice to illustrate the central limit theorem and sampling distribution respectively, and bags of bingo chips of different colors to illustrate hypothesis tests and confidence intervals for proportions. M&M® candies are a popular product for hypothesis tests for proportions [Johnson (1993), (Fricker (1996)] and other statistical exercises [Froelich & Stephenson (2012)]. Experiments have been designed using chocolate chip cookies to teach quality control [Baker (2014)] and difference between means [Magel (1998).] Martinez-Dawson (2003) documents an experiment of an interdisciplinary nature, designed expressly to demonstrate the applicability of statistics to the physical sciences.

Teachers of statistics clearly believe that experiments, *per se*, add value by supplementing problem-solving based on textbook problems and datasets. The author's position is that while textbook problems provide realistic contexts, the data are still hypothetical for students. Additionally, in the context of inferential statistics, it has been the author's experience that textbook problems are often worded in such a way that the specification of the null and alternate hypotheses is obvious. When students are asked to gather data by conducting a hands-on experiment, students must also consider what alternate specifications of null and alternate hypotheses would signify in a real-world context. Weinberg, Wiesner, and Pfaff (2010) discusses an activity "designed to help students begin to understand the concepts that underlie the T-test and confidence interval." Roy (2018) describes an experiment for hypothesis testing, confidence intervals, and tolerance intervals using boxes of '100 paper clips'. The activity is very general in context and poses "statistical" questions from a customer perspective.
The GAISE College Report recommends the use of activities, projects, and interesting datasets to help students become "actively engaged in their own learning". This is what hands-on experiments are designed to achieve. Unlike standardized lab experiments in the physical sciences, experiments designed by teachers of statistics are each unique in some respects. The question naturally arises: Do students perceive value in these data projects and experiments? Mackisack (1994) notes how, after gathering their own data sets, students "become … frustrated with the lack of information provided by textbooks, and constantly ask for

more information so that they can decide what the problem really means." This suggests that the exercise was more engaging for students compared to textbook datasets and problems. Fillebrown (1994) reports positive responses to her data project on student evaluations, a sample of which include: "The project was a good way to tie the course together", "It brought all the material together well", and "[The project] made you look at statistics in a hands-on way." The limited *qualitative* evidence that exists does seem to suggest that students find value in activities and data projects.

The primary objective of the current research was to gather *quantitative* evidence on student perceptions. Towards this end, a perception survey was designed to assess students' perceptions towards the inferential statistics experiment described in Roy (2018). This experiment was chosen because the experiment design allows students to see "their own data amongst their classmates' data", something the GAISE College Report suggests is additionally enjoyable for students.

A Likert-type survey was developed to gauge perceptions along five different dimensions that can be considered important for meaningful and effective experiments:

(i) Presentation: how well the experiment instructions were presented and explained,
(ii) Authenticity: how authentic the experiment was in students' perception,
(iii) Engagement: how engaging students felt the experiment was, as compared to textbook problems and datasets,
(iv) Learning: how much the experiment helped students understand the statistical concepts illustrated by the experiment, and
(v) Affect: students' perception about the experiment overall and its usefulness beyond the course.

The rest of the paper is organized as follows. Section 2 outlines the methodology. Findings are reported in Section 3. Section 4 concludes with some ideas for future research.

2. Experiment and Research Questions

2.1 The Experiment

The experiment was conducted in two equivalent undergraduate statistics courses. Each student in class was handed a box of '100 paper clips'; each section used a different brand. To lay the groundwork for subsequent inferences, students were initially asked to (i) identify the manufacturer's "measurable" claim, and (ii) record how many paper clips they *expected* to find in their box. They were then asked to count the *actual* number of paper clips in their individual box. Once all observations were collected and recorded, three inferential statistics questions were posed: (i) is the manufacturer's "claim", as represented by the count label, supported by the data? (ii) based on the sample, what can be inferred about the average number of paper clips packaged per box? and (iii) how many paper clips can an individual consumer expect to find in her/his box? The complete assignment is provided in Appendix I.

The activity requires all students to be involved in collecting the data, helping them understand the difference between the sample size (the number of boxes handed out) and observation (the number of paper clips in a box). Of greater significance, the specification of meaningful null and alternate hypotheses requires careful consideration and discussion. What does a count label refer to, for any packaged product? Does it refer to the average number of items contained, or the number of items in each box? What is testable based on the sample data collected? Why might this be a meaningful test? How do manufacturers (of paper clips, in this instance) calibrate their machines to comply with the count label on the package? Do they aim to fill at least what is stated on the count label, or exactly what is stated? If "at least", what is the number they aim for? Based on the observed sample mean and/or what students argue the manufacturer may be targeting, what is a reasonable set of null and alternate hypotheses? At a pedagogical level, if students argue that the manufacturer may be aiming to fill at least the number of paper clips labelled – paper clips are an inexpensive product, after all – the discussion related to "expected value" (for the hypothesis test) helps them understand why an inequality cannot belong in the null.

2.2 The Perception Survey

After the experiment had been conducted and students had made their inferences, they were asked to complete a 21-question, Likert-type survey. The survey was developed drawing on Radhakrishna (2007), Diemer *et al* (2012) and Nemoto & Beglar (2014). Twenty (20) questions related to various aspects of the experiment and the assignment instructions, with the final question asking students their gender. The 20 experiment-related questions were grouped into sets of 4 questions each, to gauge the five different dimensions of perception sought to be tested: (i) Presentation, (ii) Authenticity, (iii) Engagement, (iv) Learning, and (v) Affect. Students indicated their response on a 4-point scale from Strongly Disagree (SD) to Strongly Agree (SA).

The complete survey, organized by dimension for readers' convenience, is presented in Appendix II. The ordering of the questions on the survey completed by students was determined using a random number generator (https://www.random.org/) to generate 5 sets of 4 non-repeating integers from 1 to 20. A sample of items from each dimension is presented below.

- Presentation: I felt the purpose of the paper clips experiment was clearly explained
- Authenticity: I believe that the paper clips experiment reflected a real-life situation
- Engagement: I was more involved in this experiment-based problem than in our textbook problems
- Learning: The paper clips experiment helped me understand what a confidence interval means
- Affect: I think the paper clips experiment was a fun way to learn quality control

2.3 Research Questions

As mentioned earlier, the primary objective of this research was to assess students' general perceptions about this hands-on experiment. There were two related hypotheses that were tested as well. Was there a difference in perceptions between male and female students? It was hypothesized that there would be no difference; that students of both genders would like the experiment equally well and perceive value.
The context of the experiment is so general that all students could, in principle, relate to it in their role as consumers. Still, since participants comprised students in two courses taught by different instructors, a test of difference in perception between the two groups of students was also conducted. It was hypothesized that there would be no difference in perception.

3. Results

3.1 Profile of Respondents

A total of 40 students conducted the experiment and completed the perception survey. Five (5) surveys were partially completed and consequently had to be excluded from the study. One was an outlier and did not represent the general student perception. Of the 34 responses included in this study, 18 were from one course (Course 1 hereafter) and 16 from the other (Course 2). There were 18 female students and 16 male students.

Table 1: Composition of students by gender and course

	Gender		
	Female	Male	Total
Course 1	7	9	16
Course 2	11	7	18
Total	18	16	34

3.2 Findings From the Experiment

While the findings from the experiment are not the direct subject of this research, they impact students' perception about the experiment. It was mentioned earlier that students were initially asked how many paper clips they expected to find in their individual box. Readers may note that majority of the students said they expected to find exactly 100 paper clips in their box. Panel A of Table 2 reports what students found in their sample data:

Table 2: Results of the Experiment

	ACCO® (Course 1)	amazonbasics® (Course 2)
A. Descriptive Statistics		
Sample average	102.2	101.3
Observed variation	99 – 107	98 – 104
B. Inferential Statistics		
Hypothesis test	Null rejected	Null rejected
95% confidence interval	101-103 (*101.24 – 103.16*)	101-102 (*100.57 – 102.03*)
[95%, 99%] Tolerance interval	95-110 (*94.81 – 109.59*)	96-107 (*95.66 – 106.94*)

Note: For confidence and tolerance intervals, calculated values are in parentheses.

That the boxes contained more than 100 on average did not seem to surprise students. They reasoned that, since paper clips are an inexpensive product, manufacturers wouldn't mind "giving away" a few extra.
After discussing how manufacturers probably calibrate their machines for filling, the null and alternative hypotheses were set up as H_0: $\mu = 100$ and H_1: $\mu \neq 100$. Table 2b reports what students inferred about the population:
Students who had less than 100 paper clips in their boxes were not too pleased, especially in light of the finding that the population of boxes, on average, contained more than 100 paper clips. Though boxes are, on average, over-filled, the tolerance intervals suggest that an individual customer could find as few as 95 or 96 in her/his box while a "lucky" customer could have as many as 107 or 110. The experiment helped students understand that the count does not refer to each box. The distinction between confidence intervals and tolerance intervals was also informative for students; it helped reinforce that random variation can result in some customers getting fewer (or more) items than what is stated on the count label.

3.3 Students' Perceptions

Table 3 below presents a summary of students' responses. The items are grouped by dimension and are in the same order as in the survey presented in Appendix II.
Majority of the students agreed or strongly agreed with most statements. The statements that all students either agreed or strongly agreed with are:

1. I enjoyed the presentation of the paper clips activity (Presentation);
16. I can honestly say that collecting the data hands-on made the dataset "real" compared to textbook datasets (Authenticity);
3. I believe the experiment truly involved concepts worth learning (Learning);
20. The paper clips experiment helped me understand what a confidence interval means (Learning); and
8. I think the paper clips experiment was a fun way to learn quality control (Affect).

These responses provide evidence that students' learning experience was enhanced by the process of generating a "real" dataset based on the experiment.

Table 3: Contingency table for responses

	SD	D	A	SA	Total	Percent A+SA
Presentation						
1	0	0	19	15	34	100.0%
15	0	1	18	15	34	97.1%
18	0	2	15	17	34	94.1%
14	0	1	22	11	34	97.1%
Authenticity						
2	0	1	17	16	34	97.1%
16	0	0	12	22	34	100.0%
12	0	2	27	5	34	94.1%
13	0	2	21	11	34	94.1%
Engagement						
5	0	1	22	11	34	97.1%
9	0	1	9	24	34	97.1%
19	0	4	18	12	34	88.2%
10	0	2	14	18	34	94.1%
Learning						
3	0	0	19	15	34	100.0%
17	0	1	19	14	34	97.1%
20	0	0	19	15	34	100.0%
6	0	5	18	11	34	85.3%
Affect						
4	0	2	20	12	34	94.1%
7	4	10	20	0	34	58.8%
11	0	5	17	12	34	85.3%
8	0	0	18	16	34	100.0%

SD: Strongly Disagree, D: Disagree; A: Agree, SA: Strongly Agree

Responses to some statements (items) came as a surprise and point to the need for revision of the survey in future. Item 7, one of the four questions designed to gauge affect, stated: I will tell others about this experiment. Nearly 40 per cent of respondents (41.2%) disagreed or strongly disagreed with this statement. This was unanticipated; we had believed the appeal of this experiment was its generality, and that students would find the results intriguing enough to share with others. Since students disagreed that they would tell others about the experiment despite its contributing to their learning, we propose to replace this item with "I will remember this experiment even after this semester is over" in a revised version of the survey.

Means and standard deviations for the overall survey as well as for each subscale are reported in Table 4 below (on a 4-point scale and with 4 items in each set, the highest possible score for each category is 16). Cronbach's alpha (reliability test) for the survey was 0.888.

Table 4: Summary Statistics

Dimension	Mean	Standard deviation
Overall	66.59	6.43
Presentation	13.59	1.74
Authenticity	13.44	1.38
Engagement	13.68	1.65
Learning	13.44	1.44

Affect	12.44	1.58

In students' perception, the experiment scored highest on Engagement and Presentation. Students perceived the experiment as authentic and indicated that it enhanced their learning.

Of the five dimensions along which students' perception was measured, the average score for Affect is found to be the lowest. One item in this set, I will tell others about this experiment, was noted earlier because of the relatively high negative response. This item had the lowest mean (2.47) and the lowest correlation (0.200) with other items on the survey. (Removing this item from the analysis improved the reliability score of the survey to a Cronbach's alpha of 0.894.) The other survey item exhibiting weak correlation with the rest of the survey, also in the Affect category, was: The experiment piqued my curiosity about how products are counted for packaging. This item had a correlation of only 0.221 with the other survey items. In future use, we propose to reword these two statements in the survey.

3.4 No Gender Difference in Perceptions

In Table 5 below, we report means and standard deviations of component scores by gender. On a 4-point scale, and with 4 items in each category, the highest possible score is 16. The last column presents t-statistics for difference in means between the two gender subgroups assuming equal variances (the F-test indicated equal variances for every component at $\alpha = 0.05$).

Table 5: Means and Standard deviations of scores by gender

	Female students (n = 18)		*Male students (n = 16)*		
Component	**Mean**	**Std. dev.**	**Mean**	**Std. dev.**	**t-stat**
Overall	66.17	5.49	67.06	7.51	-0.40
Presentation	13.61	1.54	13.56	2.00	0.08*
Authenticity	13.33	1.19	13.56	1.59	-0.48
Engagement	13.56	1.46	13.81	1.87	-0.45
Learning	13.28	1.41	13.63	1.50	-0.70
Affect	12.39	1.50	12.50	1.71	-0.20

<u>Note</u>*: * denotes significance at 10%*

Overall, male students had a more positive perception of the experiment than their female counterparts, but the difference is not statistically significant. Component-wise, they scored the experiment higher than female students on all but presentation. The difference in perceptions about presentation, while statistically significant, is probably not of practical significance.

Male students found the experiment more engaging overall when all four items in the Engagement category are taken together. It is therefore interesting to note that the highest item score (mean of 3.7 on a 4-point scale) for female students was on the item, "I was more involved in this experiment-based problem than in our textbook problems", included in the engagement component of the survey.

Among male students, two items were tied for highest score (3.69 average). One item was related to engagement: I was more involved in this experiment-based problem than in our textbook problems. The other related to authenticity: I can honestly say that collecting the dataset hands-on made the dataset "real" compared to textbook datasets.

Taken together, students' responses indicate that the experiment helped supplement textbook data sets and make the context "real". This is the most rewarding finding of the study. Teachers of statistics will agree that providing students the experience of collecting and analyzing real data sets is “the” *prima facie* case for incorporating experiments in statistics classes. While students perceived value differently, with female students finding value in indicators different from those indicated by males, the overall perception was positive and validates the use of experiments in statistics classes.

3.5 Difference in Perceptions Between Courses

An unexpected an intriguing finding of our study was the significantly different perception of the two groups of students (two courses). Table 6 below presents results of a t-test for difference in means, assuming equal variances (the F-test indicated equal variances for every component at $\alpha = 0.05$). Students in course 1 scored the experiment better overall, and on the three components of presentation, engagement, and learning. There was no significant difference in perception along the two dimensions of authenticity and affect.

Table 6: Means and Standard deviations of scores by course

	Course 1 (n = 16)		*Course 2 (n = 18)*		
Component	**Mean**	**Std. dev.**	**Mean**	**Std. dev.**	**t-stat**
Overall	68.75	6.92	64.67	5.45	1.92**
Presentation	14.06	1.77	13.17	1.65	1.53*
Authenticity	13.75	1.34	13.17	1.38	1.25
Engagement	14.31	1.62	13.11	1.49	2.25**
Learning	13.88	1.54	13.06	1.26	1.70*
Affect	12.75	1.77	12.17	1.38	1.08

*Note: ** denotes significance at 5% and * at 10% level of significance*

The difference in perception can be attributed to two possible factors. There was a difference in the timing of the activity, *ie*. the placement of the experiment in the lesson schedule in the two courses. In course 1, the experiment was conducted immediately after the discussion on confidence intervals and hypothesis testing had been completed, before students took a test on this unit. In course 2, the experiment was conducted on the last day of lessons before final exam, with a resultant time lag between the topics directly related to the activity and the experiment itself. If this is indeed a factor that affects relevance of an activity and perceptions thereof, it underscores the importance of timing activities to keep them relevant and current to the topic at hand (akin to lectures and labs in the physical sciences).

A second possible factor is the fit of an activity of this nature in the overall lesson design and teaching style. The courses were taught by two different instructors with their unique teaching styles, lesson plans, and data projects; this was the only activity that was common to the two courses. It is possible that the activity was a better fit overall in one course than in the other.

Course-wise, the following items received the highest average scores overall:

1. Both courses: I can honestly say that collecting the data hands-on made the dataset “real” compared to textbook datasets (Authenticity);
2. Course 1: Compared to other class activities, the paper clips experiment was far more engaging (Engagement);
3. Course 2: I was more involved in this experiment-based problem than in our textbook problems (Engagement).

Despite some differences in perception, both sections validated the use of this experiment to supplement textbook examples.

4. Conclusions and Direction for future research

The research reported in this paper is an extension of Roy (2018) that describes a hands-on inferential statistics experiment requiring students to collect data and answer questions about a packaged product. To assess whether students perceive value in such an experiment, a perception survey was designed to gauge student feedback on different aspects of the activity. The evidence suggests that students indeed perceive positive benefits from such experiments. No gender difference in student perceptions was discerned. An interesting but unexpected finding was a significant difference in perception between the two sets of students that conducted the experiment. This could be partly attributable to the timing of the assignment. How well an assignment of this nature fits into the overall structure of lessons in the course, may also be an important factor. The contribution of this research is the formal analysis of value derived by students from hands-on statistics experiments.

There are two directions in which we propose to extend this research in future. The first is assessing how well the positive perception correlates with a formal assessment of students' understanding of the concepts, on subsequent formative or summative exams. Weinberg, Weisman, and Pfaff (2010) noted how their hands-on activity "generally failed to help students develop a good understanding of the underlying statistical concepts", despite students having "a positive reaction to the modules." A similar analysis for the experiment used in this study would help ascertain its value in enhancing students' understanding.

Another extension would be to test whether incorporating experiments into statistics lessons improves students' attitudes towards the subject. Low and declining attitudes towards statistics appears to be a common finding in many studies [Gal and Ginsburg (1994), Schau and Emmioglu (2012)]. While documenting the timeline of attitudes is informative [Kerby and Wroughton (2017), it may be of greater consequence and benefit to identify what educational practices can be put in place to not only stem this decline but help students develop a positive attitude towards statistics. Based on the positive perception students showed towards this experiment, we believe it may be worthwhile to design and incorporate experiments at various points in the semester, illustrating different statistical concepts, to keep students actively engaged and interested in the subject.

Acknowledgements

This paper was presented in a speed presentation format and corresponding poster session at the 2018 Joint Statistical Meetings. I thank attendees for their positive comments.

References

Baker, A. (2014). Teaching Quality Control with Chocolate Chip Cookies. *Teaching Statistics*, 36(1), 2-6.

Diemer, T. T., Fernandez, E., and Streepy, J. W. (2012). Student Perceptions of Classroom Engagement and Learning using iPads. *Journal of Teaching and Learning with Technology*, 1(2), 13-25.

Fillebrown, S. (1994). Using Projects in an Elementary Statistics Course for Non-Science Majors. *Journal of Statistics Education*, 2(2). DOI: 10.1080/10691898.1994.11910470

Fricker, R.D. Jr. (1996). Yikes! Where Did All the Blue M&M's Go? The Mysterious Case of the Blue M&M's®. *Chance*, 9(4), 19-22.

Froelich, A. G. and Stephenson, W. R. (2013). How Much Do M&M's Weigh? *Teaching Statistics*, 35: 14-20. doi:10.1111/j.1467-9639.2012.00515.x

GAISE College Report ASA Revision Committee (2016). Guidelines for Assessment and Instruction in Statistics Education College Report 2016.

Gal, I., and Ginsburg, L. (1994). The Role of Beliefs and Attitudes in Learning Statistics: Towards an Assessment Framework. *Journal of Statistics Education*, 2(2). DOI: 10.1080/10691898.1994.11910471

Johnson, R. W. (1993). Testing Colour Proportions of M&Ms. *Teaching Statistics*, 15: 2–4. doi:10.1111/j.1467-9639.1993.tb00243.x

Kerby, A. and Wroughton, J. R. (2017). When Do Students' Attitudes Change? Investigating Students' Attitudes at Midterm. *Statistics Education Research Journal*, 16(2), 476-486.

Lind, D., Marchal, W., and Wathen, S. (2018). *Statistical Techniques in Business and Economics*. McGraw-Hill.

Mackisack, M. (1994). What Is the Use of Experiments Conducted by Statistics Students? *Journal of Statistics Education*, 2(1). DOI: 10.1080/10691898.1994.11910461

Magel, R. C. (1998). Testing for Differences Between Two Brands of Cookies. *Teaching Statistics,* 20(3), 81-83.

Martinez-Dawson, R. (2003). Incorporating Laboratory Experiments in an Introductory Statistics Course. *Journal of Statistics Education*, 11(1). DOI: 10.1080/10691898.2003.11910697

Nemoto, T., & Beglar, D. (2014). Developing Likert-scale questionnaires. In N. Sonda & A. Krause (Eds.), *JALT2013 Conference Proceedings*. Tokyo: JALT.

Pfaff, T. P., and Weinberg, A. (2009). Do hands-on activities increase student understanding?: A case study. *Journal of Statistics Education*, 17(3). DOI: 10.1080/10691898.2009.11889536

Radhakrishna, R. B. (2007). Tips for Developing and Testing Questionnaires / Instruments. *Journal of Extension*, 45(1).

Roy, S. (2018). Inferential Statistics Made Fun: Stories that Boxes of '100 Paper Clips' can Tell. *Teaching Statistics* [online]. Retrieved from https://doi.org/10.1111/test.12168

Schau, C., and Emmioglu, E. (2012). Do Introductory Statistics Courses in the United States Improve Students' Attitudes? *Statistics Education Research Journal*, 11(2), 86-94.

Weinberg, A., Weisner, A. and Pfaff, T. J. (2010). Using Informal Inferential Reasoning to Develop Formal Concepts: Analyzing an Activity. Journal of Statistics Education, 18(2). DOI: 10.1080/10691898.2010.11889494

Appendix I

Inferential Statistics Experiment

Purpose: Apply your (students') understanding of inference to a real situation.
Learning outcomes: Hypothesis test, Confidence interval, Tolerance interval.
Completion: Data *collection* - Individual; Data *analysis and interpretation of findings* - Group.

Introduction

In this hands-on experiment, you will analyze data collected from a sample of [*insert brand name*] paper clips. Each of you will be handed one box of paper clips. The sample size is the total number of boxes handed out to the class for the experiment.

Instructions

I. Understand the "claim"

Once you are handed a box, identify the manufacturer's "claim" stated on the box.
Also record how many paper clips you *expect* to find in your box.

The manufacturer's "claim" is ________________________________.

I *expect* my box to contain ___________________________________.

II. Collect data

Please count the number of paper clips in your box. Arrange your clips in batches of five. This will help avoid counting errors. It will also make it easier for you to double-count for accuracy.

Record the number of paper clips on this sheet. Also enter it into the single Excel data file being created for this experiment.

The *actual* number of paper clips contained in my box is ______________.

The total number of observations (*ie*. boxes of paper clips) in our data set is _____________.

Note for instructor: Please project the Excel file onto the screen so students can see the data as it gets populated. Then, please provide students access to the data file if you want them to do the analysis. Alternatively, they can state what summary statistics they need, and you can demonstrate the calculations.

III. Analyze the data and make inferences

a. <u>Hypothesis test</u>: is the manufacturer's claim supported by the data?

At a 5% significance level, test the manufacturer's "claim" stated on the box. Provide all five steps, starting with the null and alternate hypotheses and ending with your conclusion.

b. <u>Confidence interval</u>: based on the sample data, what can you infer about the "true" average number of paper clips contained per box?

Construct the 95% confidence level for the "true" mean number of paper clips per box. Is the manufacturer's "claim" contained in your confidence level?

c. <u>Tolerance interval</u>: how many paper clips can an individual customer expect to find in her/his box?

The actual number of paper clips contained in each box deviates from the average, as you saw at the data collection stage. Construct the [95%, 99%] tolerance interval to infer, with 95% confidence, the actual number of paper clips contained in 99% of boxes. Use the tolerance interval applet at http://statpages.info/tolintvl.html. Complete the following sentence.

Based on the _____________ interval, we can be ______ confident that ______ of boxes contain between __________________ paper clips.

Appendix II

Survey to Gauge Students' Perception of Experiment

Presentation
1. I enjoyed the presentation of the paper clips activity
15. I felt the purpose of the paper clips experiment was clearly explained
18. I believe the instructions for the activity were presented in an organized manner
14. I learned a lot from the presentation of the paper clip activity

Authenticity
2. I believe that the paper clips experiment reflected a real-life situation
16. I can honestly say that collecting the data hands-on made the dataset "real" compared to textbook datasets
12. I could relate to the questions that the paper clips experiment was designed to answer
13. The context of the experiment seemed very "real life" to me

Engagement
5. I was intrigued by the results of the paper clips experiment
9. I was more involved in this experiment-based problem than in our textbook problems
19. I was far more captivated by statistics following this paper clips experiment
10. Compared to other class activities, the paper clips experiment was far more engaging

Learning
3. I believe the experiment truly involved concepts worth learning
17. The paper clips experiment helped me understand hypothesis testing better
20. The paper clips experiment helped me understand what a confidence interval means
6. I have gained knowledge from the experiment that will be useful beyond this course

Affect
4. I really liked the paper clip experiment
7. I will tell others about the paper clips experiment
11. The experiment piqued my curiosity about how products are counted for packaging
8. I think the paper clips experiment was a fun way to learn quality control

Finally, what is your gender?

☐ **Female** ☐ **Male**