TEACHING STATISTICS WITH GEOGEBRA *

Tuyetdong Phan-Yamada & Sui Wing Man

California State University, Los Angeles

Abstract

In this article, the author discusses a statistics project that examines the ability of participants to accurately estimate the location of midpoints of various line segments from a GeoGebra sketch. Students explore differences in estimating accuracy for various types of lines among different subgroups of subjects. Keywords: GeoGebra; statistics; data analysis

1 Preliminaries

GeoGebra is well-known for its dynamic graphing tools for geometry, trigonometry, and calculus (Quinlan, 2013). But did you know that the software also includes powerful tools for statistical analysis? For instance, with GeoGebra, students can create statistical graphs, conduct hypothesis tests, and run statistical simulations of real-world phenomena. In this paper, we highlight the use of GeoGebra as a tool to collect and analyze data from a geometric estimation activity. The activity—the *Line Segment Midpoint Challenge*—connects geometry with descriptive and inferential statistics to create rich learning opportunities for students.

2 Introducing the Project

Locating the midpoint of line segments is the basis of the *Line Segment Midpoint Challenge*. We introduce the project to our students by providing them with a sketch similiar to that shown in Figure 1. After providing students with a few minutes to explore the sketch and its data collection capabilities, we ask the students to brainstorm questions about data they might collect with the software. Popular questions include the following:

- Is it easier for people to locate the midpoint of a horizontal segment or an oblique segment?
- Do people get better estimating the position of midpoints of line segments over time? Can people learn to get better at guessing the position of line segments? If so, does this ability "stick" (i.e., does it last over time–say several weeks)?
- Are females better equipped to locate the midpoint of segments?
- Is one's ability to locate the midpoint of a segment influenced by factors such as gender, age, or academic major?

*All sketches in this paper are available in a GeoGebraBook at `https://tinyurl.com/challenge-sketchbook`. Additional materials are available at `https://sites.google.com/site/phanyamada/Home/conference`

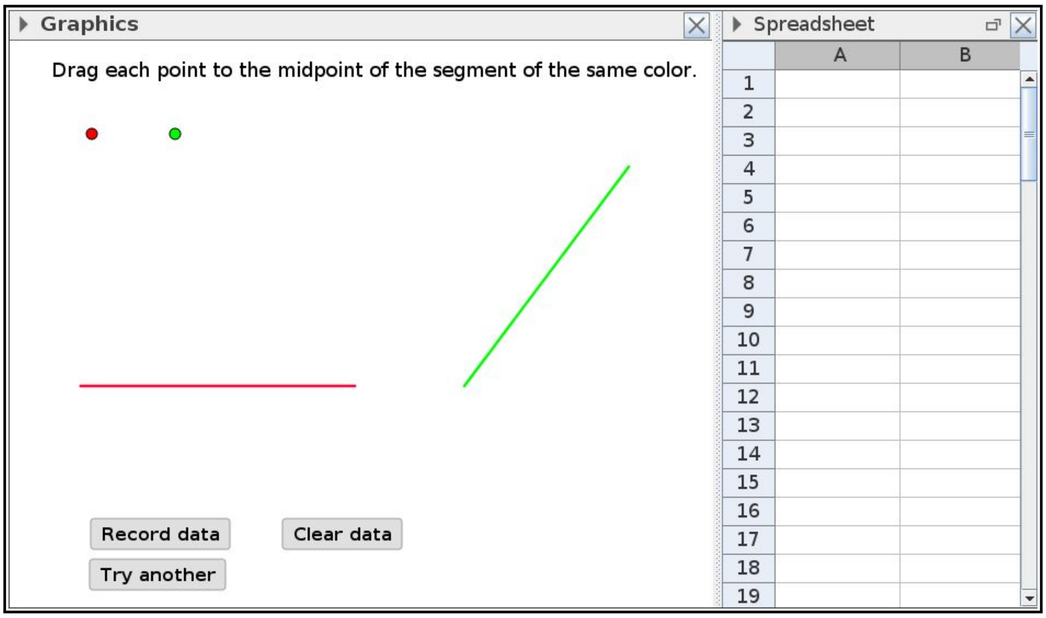


Figure 1. Initial *Line Segment Midpoint Challenge* sketch (available on-line at `https://tinyurl.com/midpoint-challenge`)

Ultimately, students will capture participant estimates using the GeoGebra sketch, then conduct hypothesis tests with the data collected within the spreadsheet. As students use GeoGebra to answer their own questions, they learn more than possible using textbook problems alone.

In this paper, we illustrate methods that one group of students used to answer a popular hypothesis—namely, that participants are able to estimate the location of the midpoint more accurately with horizontal segments than with oblique ones in a single sitting. You are encouraged to download and modify the sketches provided within this paper and explore them with your students.

3 Data Collection

Our students collect participant data using a sketch similar to the one highlighted in Figure 1 (available on-line at `https://tinyurl.com/midpoint-challenge`). Subjects drag each colored point to the midpoint of the like-colored segment (e.g., they drag the red point to the midpoint of the red segment). Clicking the “Record Data” button precipitates three actions: (a) the distance between each estimate and the actual midpoint is displayed on the screen for users to consider; (b) these distances are recorded to the first two columns of the spreadsheet in the right-most window of the sketch; and (c) the estimates are frozen—users can no longer move the red or green points. Clicking on the “Try Another” button removes the distance text from the screen while returning the points to their initial positions. The same participant can estimate the location of the midpoints again, or a new subject can try. Figure 2 highlights collected data gathered from 40 participants. The sketch provides students with an appealing, portable data collection tool. For instance, one of our students collected data from participants at a local shopping mall. Many people came to try and were excited to see how close they could get to the actual midpoint. Imagine collecting the same data with paper and pencil! Without question, students would have a more difficult time recruiting participants. Moreover, measuring error would become a much more time consuming enterprise. With the tool, our students were able to gather data from many subjects quickly.

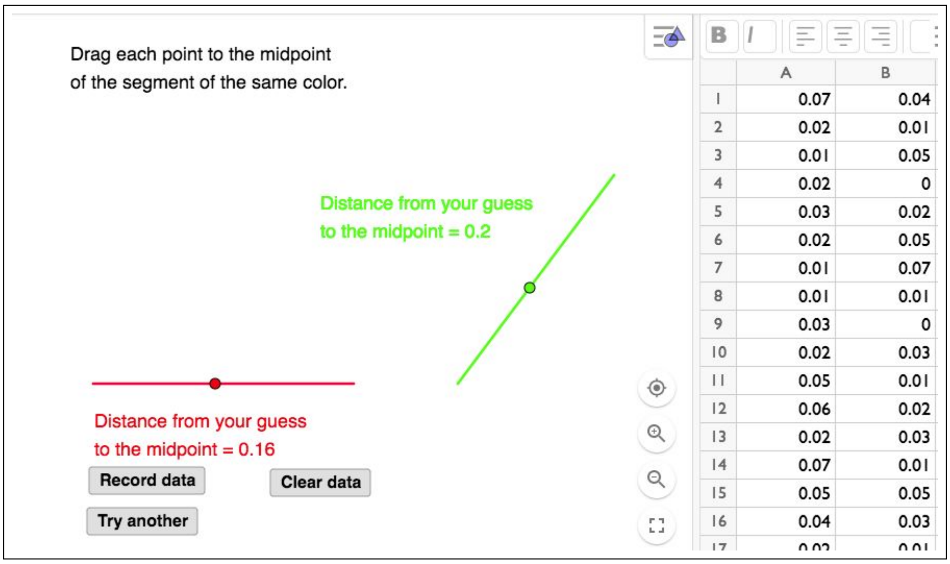


Figure 2. *Line Segment Midpoint Challenge* sketch with data generated from 40 participants (available at `https://tinyurl.com/challenge-and-data`).

4 Analyzing the Collected Data

4.1 Correlation and Outliers

A first task that students typically do with their data is examine whether a correlation exists between the data for the horizontal and the slant segment. Let's explore some sample data at `https://tinyurl.com/challenge-and-data`. As Figure 3 suggests, students highlight columns A and B, then select the "Two Variable Regression Analysis" by clicking the histogram button from the top left (why don't you try this, too?).

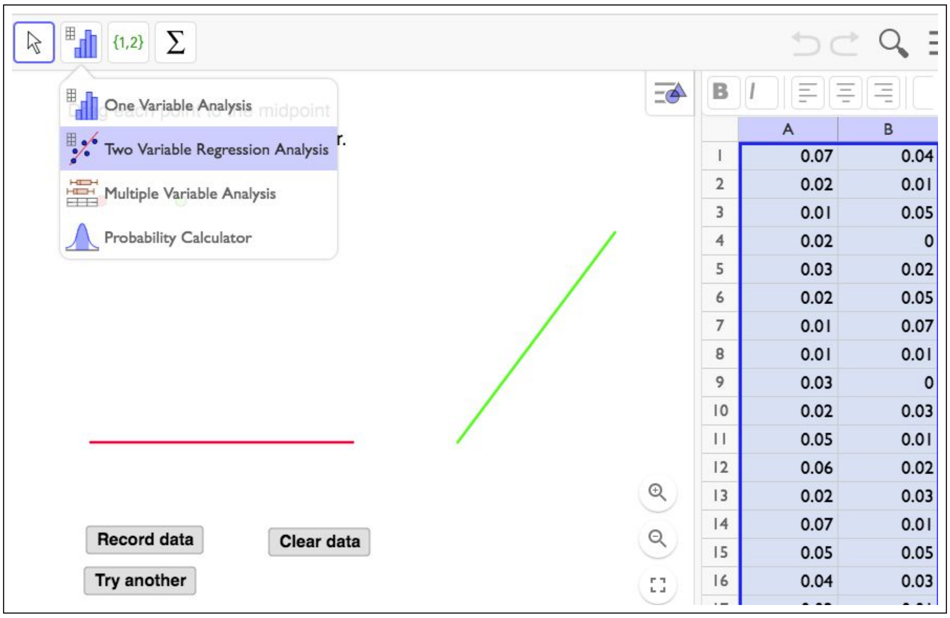


Figure 3. Selecting the **Two Variable Regression Analysis** tool in GeoGebra.

Figure 4 provides a scatter plot of the slant segment midpoint error with respect to the horizontal segment midpoint error along with a linear regression model.

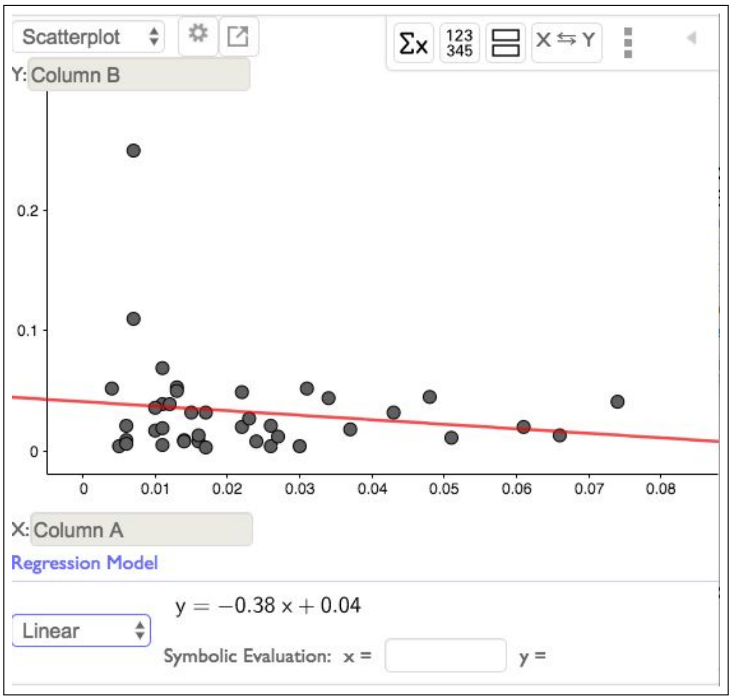


Figure 4. Exploring correlation with GeoGebra.

Notice that Figure 4 includes an outlier. Let's do some analysis in GeoGebra to further explore this. At this point, the Graphics and Spreadsheet views of the sketch are no longer needed. Uncheck these to toggle them off, as shown in Figure 5.

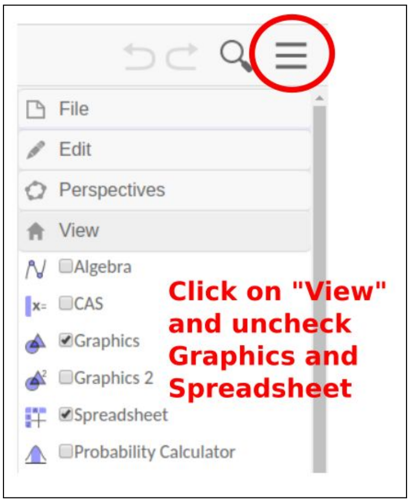


Figure 5. Steps to toggle off Graphics and Spreadsheet views in GeoGebra online app.

Next, click on the **Show Statistics** and **Show Data** tools in the Data Analysis toolbar (as shown in Figure 6). This allows users to select which data to plot and to explore various statistical measures related to the collected data.

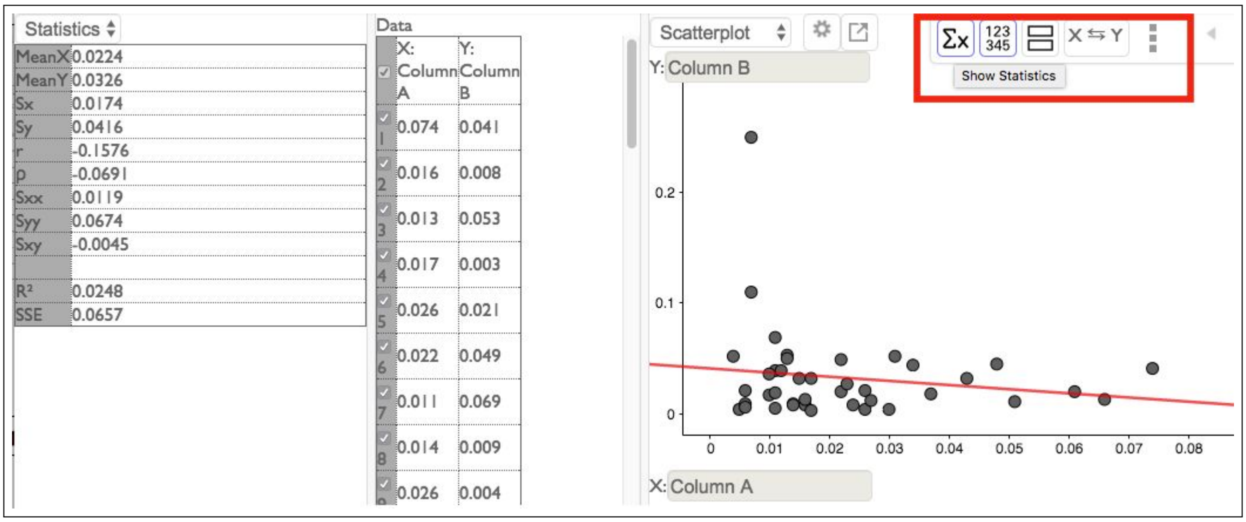


Figure 6. Show Statistics and **Show Data** tools within the GeoGebra's Data Analysis application (available at `https://tinyurl.com/analyze-midpoint-data`).

As Figure 6 suggests, the correlation coefficient for the collected data, r, is approximately -0.1576, suggesting little or no association between the variables. Let's remove the outliers by unchecking data element #33, the item with the largest y-value (i.e., 0.25). The scatter plot will refresh to exclude ordered-pair 33. As Figure 7 suggests, the correlation coefficient becomes $r = -0.0685$, which indicates that there is almost no correlation between the error of locating the midpoint of a horizontal segment and the error of locating the midpoint of an oblique segment.

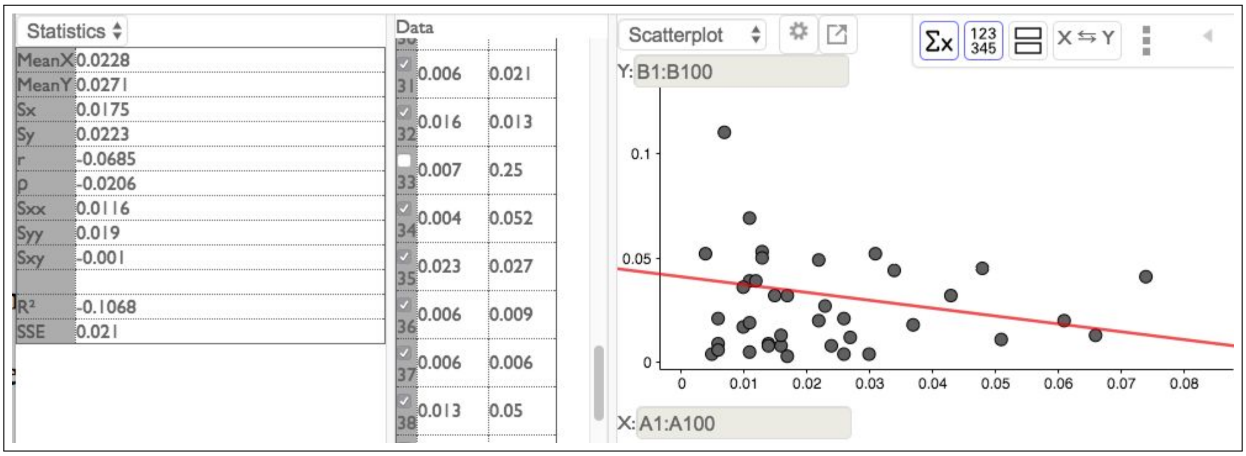


Figure 7. Statistical measures are dynamically updated as students unselect outliers.

We've written a mini project with more details on this topic. For more ideas, download the project, MATH 1090 PROJECT Midpoint.pdf, at `https://sites.google.com/site/phanyamada/Home/conference`.

4.2 Confidence Intervals

In addition to exploring correlation with with our students, we have them find the confidence interval for the error of estimating the midpoint of a horizontal or a slant segment using GeoGebra. Let's do similar work together. Begin by navigating to our original sketch with data, provided at `https://tinyurl.com/challenge-and-data`. Highlight the data in the left-most column, A, and select **"One Variable Analysis"** by clicking the histogram button from the top left. As we did in our earlier work, hide the Geometry and Statistics views and click on the **Show Statistics** and **Show Data** tools in the Data Analysis toolbar (as shown in Figure 6). After that, we click the statistics drop down menu from the **Show Statistics** pane and choose **"t estimate of a mean"** as suggested in Figure 8.

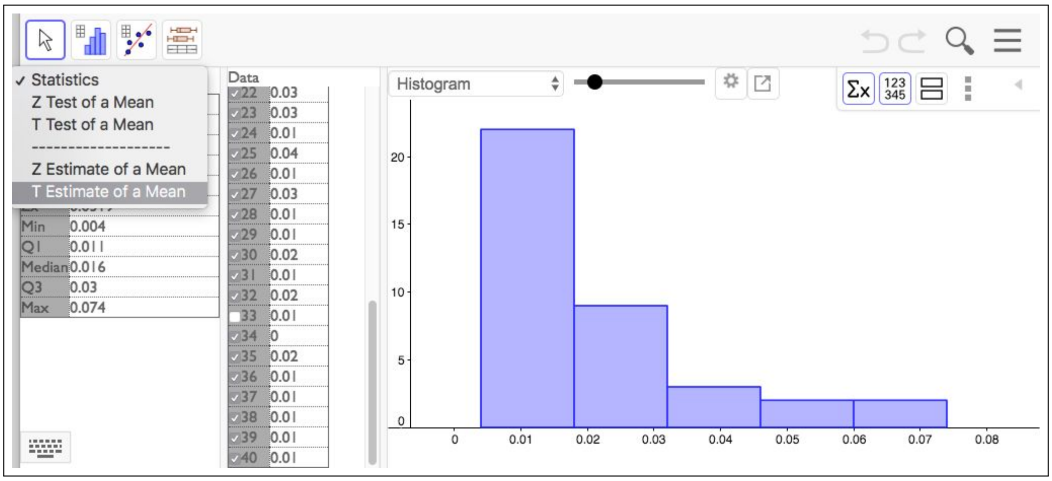


Figure 8. (Left) Selecting the t estimate of a mean function; (Right) A histogram of estimation errors for the horizontal segment (with outlier removed).

Lastly, enter a desired confidence level and press Enter. GeoGebra will give the statistics of the data set including the graph. Try changing the width of the histogram by moving the slider on the right side of the histogram button or choosing different graphs by clicking the histogram drop down menu, as suggested in Figure 9.

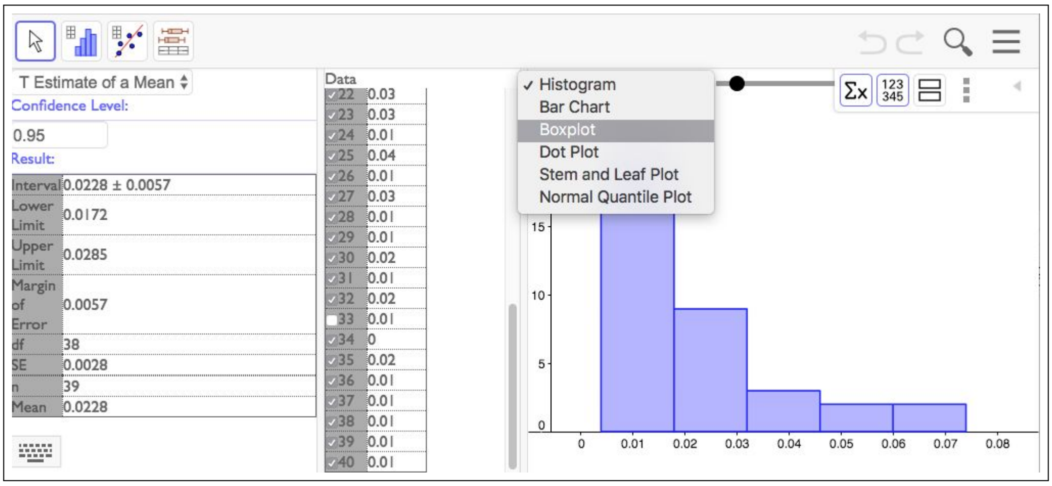


Figure 9. (Left) Results from t estimate of the mean with a confidence level of 0.95; (Right) Choosing different graphs from the histogram drop down menu.

Clicking on the gear to the right of the histograph slider reveals various options, including the choice to show the frequency table. As Figure 10 suggests, the confidence interval for this data set is $(0.0172, 0.0285)$.

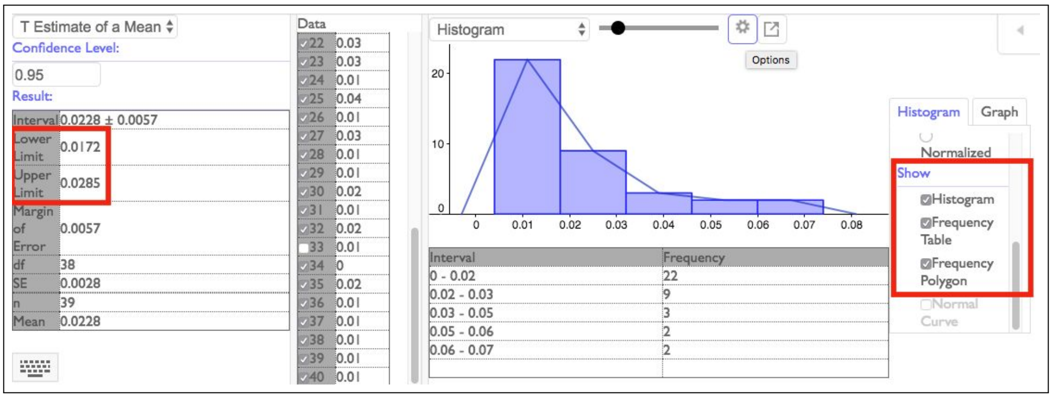


Figure 10. ((Left) Lower and upper limits of confidence interval; (Right) Histogram settings.

Once each group of students produces a confidence interval, the intervals can be plotted together using the sketch provided at `https://tinyurl.com/challenge-intervals`. We start by entering the lower and upper limit of the confidence interval of each group to the spreadsheet. If the range of the data set is too small (or too big), multiply (or divide) data by a power of 10 to generate a graph that is more readable. For instance, for our data the range was too small, from 0.0102 to 0.039. So our students multiplied by 100 before entering upper and lower bounds and means into the sketch. Lastly, we entered the number of student groups entering data (in our case, $n = 20$). As Figure 11 illustrates, confidence intervals are graphed with the first row appearing on the bottom of the graph and the last row appearing on the top.

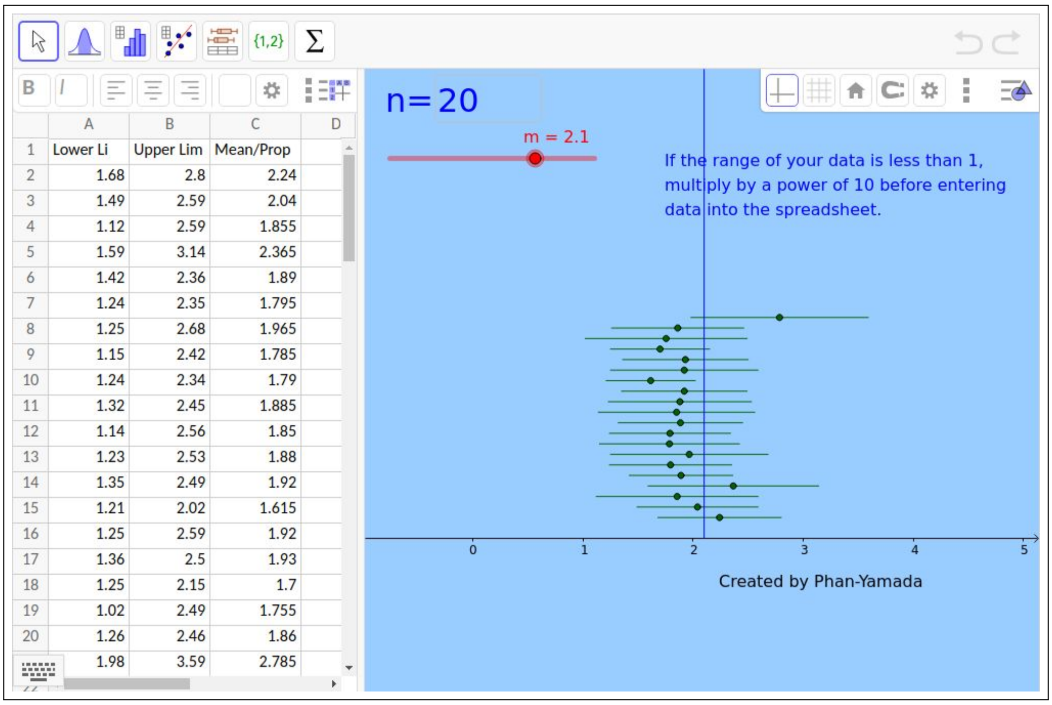


Figure 11. Confidence intervals from the t Estimate of the mean for 20 different student groups (available at `https://tinyurl.com/confidence-exploration`).

For instance, the confidence interval $(1.68, 2.8)$ is represented by the lowest green bar. Use the red slider m to move the blue line. Figure 11 suggests that 19 out of 20 (or 95%) of the different confidence intervals contain the value $m = 2.1$, reinforcing the idea that with a 95% confidence level, we expect about 19 out of 20 confidence intervals to contain the true value of error 0.021 when estimating the midpoint of a horizontal segment.

5 Hypothesis Tests

5.1 t-test of Mean error

First, we explore hypothesis tests with GeoGebra by testing the claim that the mean of error between the estimated and actual midpoint of a horizontal segment is 0.021. We return to our original sketch with data, provided at `https://tinyurl.com/challenge-and-data`. With the left-most column highlighted (i.e., Column A), select **"One Variable Analysis,"** as illustrated in Figure 12.

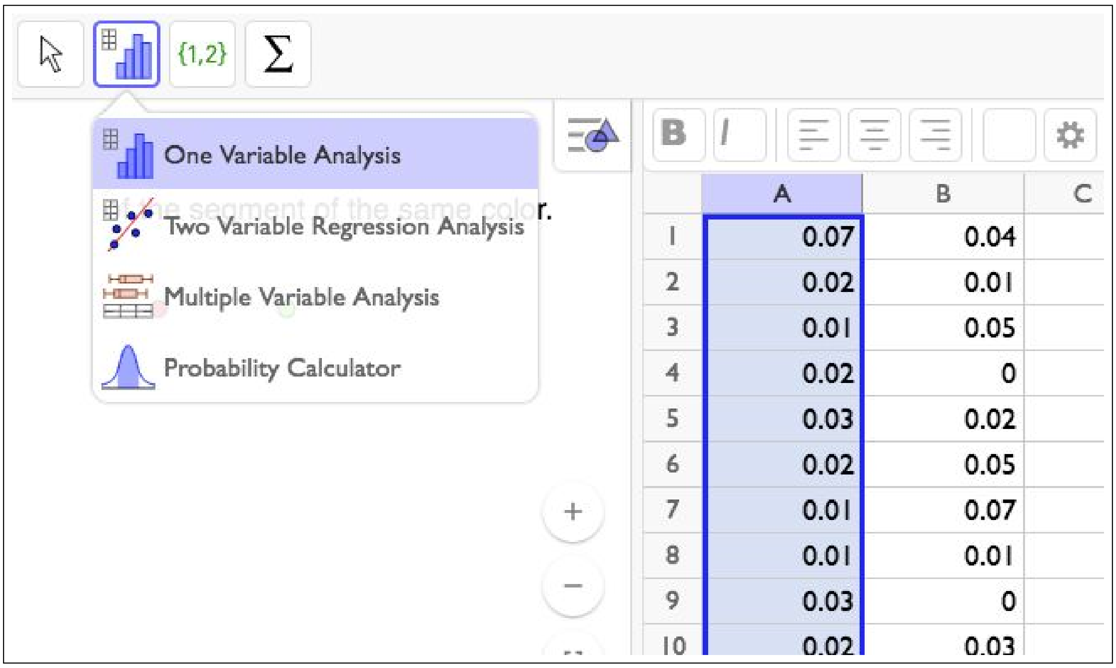


Figure 12. Highlighting mean error data and selecting **"One Variable Analysis"** tool.

Then we hide Graphics and Spreadsheet views, click the **"Show Statistics"** tool, and choose **"t-Test of a Mean"** from the drop down menu. Enter the value for the Null Hypothesis as $\mu = 2.1$ and the Alternative Hypothesis as $\mu =\neq 2.1$. Figure 13 suggests these steps and shows that the p-value is approximately 0.6081. Since this is greater than the confidence level 0.05, we do not reject the null hypothesis $\mu = 0.021$. Furthermore, students can also check for normality by selecting **"Normal Quantile Plot"** from the **"Histogram"** drop down menu, which suggests normality in this case.

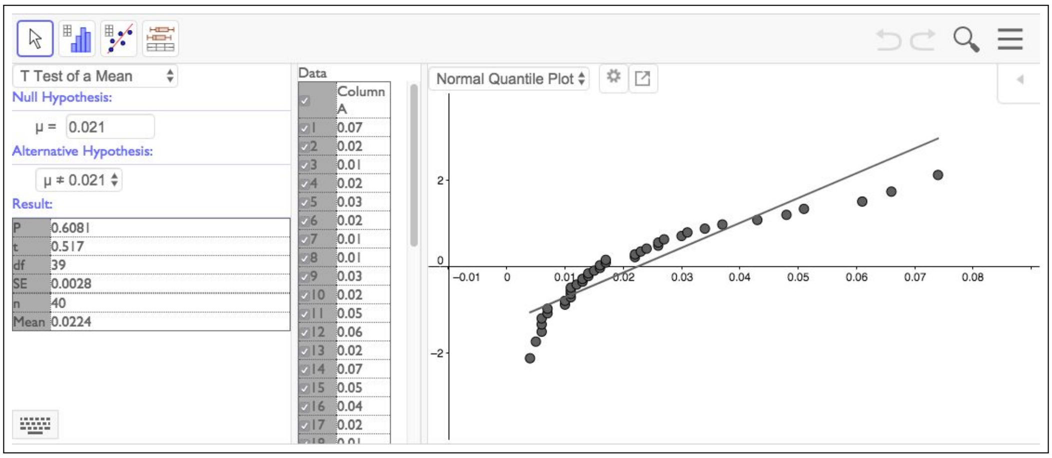
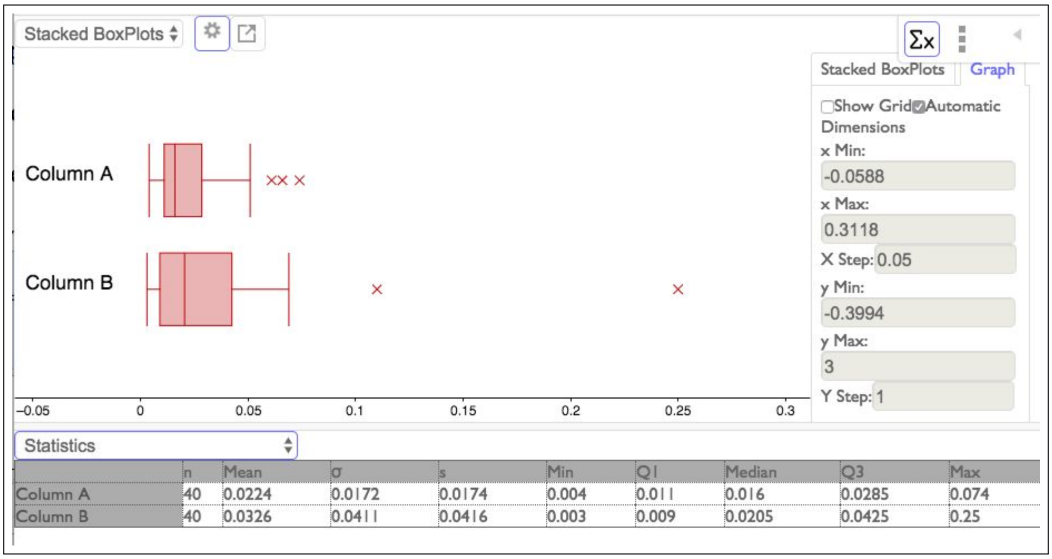


Figure 13. t-Test of a mean and normal quantile plot.

5.2 *Comparing mean error for horizontal and oblique segments*

Lastly, we conduct a t-test for the paired differences to test the claim that people can locate the midpoint more easily on a horizontal segment than on a slant segment. This means the error of the horizontal segment is smaller than the error of the slant segment, or equivalently, the mean of the difference of the horizontal segment error and the slant segment error is less than 0. We will compare the error of locating the midpoint of a horizontal segment and that of a slant segment of every participant. Again, we use our original sketch with data, provided at `https://tinyurl.com/challenge-and-data`, this time highlighting data in columns A and B (column A contains horizontal segment data; column B, oblique data). Select **"Multiple Variable Analysis"** by clicking the histogram button from the top left. This will display the boxplot of these two sample sets with their outliers marked with X. As we did in our earlier work, hide the Geometry and Statistics views and click on the **Show Statistics** tool in the Data Analysis toolbar as shown in Figure 14.



	n	Mean	σ	s	Min	Q1	Median	Q3	Max
Column A	40	0.0224	0.0172	0.0174	0.004	0.011	0.016	0.0285	0.074
Column B	40	0.0326	0.0411	0.0416	0.003	0.009	0.0205	0.0425	0.25

Figure 14. Multivariable analysis of horizontal and oblique segment midpoint estimation error.

Now we perform the hypothesis test. We click the Σx button followed by clicking the **"Statistics"** drop down menus to select **"t-Test, Paired Differences"** and change the option for **"Alternative Hypothesis"** to **"Mean Difference $<$ 0."** After this, all the statistics are shown as in Figure 15.

T Test, Paired Differences

Null Hypothesis: Mean Difference = 0

Alternative Hypothesis: Mean Difference < 0

	Mean	s	n
Sample 1 Column A	0.0224	0.0174	40
Sample 2 Column B	0.0326	0.0416	40
Differences	-0.0102	0.0475	40

Mean Difference	P	t	SE	df
-0.0102	0.0913	-1.3567	0.0075	39

Figure 15. Paired t-test results

Since the p-value 0.0913 is more than 0.05, we do not reject the null hypothesis. There is not enough evidence to support the claim that people locate the midpoint of the horizontal segment better than the slant segment.

6 In Conclusion

By implementing Geogebra to this project, we eliminate possible confounding factors such as carelessness of the participants when they mark the midpoint, or the lack of skills and instruments to make accurate measurements on the behalf of the students. This GeoGebra app takes care of all these problems, and on top of that, it makes both students and participants excited to find out the result.

Once students enter their data on a GeoGebra spreadsheet, all calculations can be done quickly. And with its graphing and statistics features, students can analyze their data much more easily. Most students in an introductory statistics course are not ready to work on huge amount of data, tedious calculations or complicated coding, but they need experience in doing research and analyzing data. With GeoGebra, instructors can teach them to do these tasks easily. When students become more efficient in reading and writing statistics papers, they can explore more into the field through coding in other computer languages (such as R, Octave or MATLAB). Readers can find many GeoGebra tools and projects for math courses from Algebra up to Differential Equations at the website (Phan-Yamada, 2018) .

References

GeoGebra (n.d.) *Statistics Commands*. Retrieved from `http://wiki.geogebra.org/en/Statistics_Commands`.

Phan-Yamada, T. (n.d.) *Statistics resources*. Retrieved from `https://sites.google.com/site/phanyamada/Home/teaching/statistics`

Phan-Yamada, T. (2018). *Author's Homepage*. Retrieved from `https://sites.google.com/site/phanyamada/Home`

Quinlan, J. (2013). GeoGebra as a frontend to generating graphics for LaTeX. *North American GeoGebra Journal, 2*(**1**), 37–42.

Tuyetdong Phan-Yamada, `tphanyamada@yahoo.com`, is a lecturer at the Department of Mathematics, CSU Los Angeles. She enjoys building interactive graphical illustrations with GeoGebra, which she integrates into her lesson plans in Trigonometry, Geometry, Statistics, Calculus, and other courses.

Sui Wing Man, `sman@calstatela.edu`, teaches math at California State University Los Angeles. In his journey as a math educator, he pursues methods to help students think like an expert and learn as a community. He is currently taking part in developing a flipped Calculus course at CSU.