



Handwriting and Gender: A multi-use data set

Sean Bradley

Clarke University

Journal of Statistics Education Volume 23, Number 1 (2015),
www.amstat.org/publications/jse/v23n1/bradley.pdf

Copyright © 2015 by Sean Bradley, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the author and advance notification of the editor.

Key Words: Handwriting analysis; Gender; Data project.

Abstract

Can individuals guess the gender of a writer based on a sample of his or her handwriting? We administer an electronic survey twice to the same individuals to find out. The resulting data set is interesting to students, rich enough to be amenable to a wide array of activities, and open to a variety of exploratory tasks for statistics students and teachers.

1. Introduction

The *Guidelines for Assessment and Instruction in Statistics Education* contains recommendations for statistics teachers, among these recommendations are fostering active learning in the classroom, using real data, helping students learn to utilize technology for analyzing the data, and facilitating the development of statistical thinking (Aliaga et al. 2012). Much has been written of the use of extended projects (Bailey, Spence, and Sinn 2013) as a way to achieve some of these goals for students. In this article, we propose the extended use of a rich and interesting student-centered data set.

The data come from a handwriting recognition survey that can be administered easily using paper and pencil, although we prefer an electronic method. We will say more about the survey itself in a later section. Teachers can use the data provided by the author, or more ideally, have their own students generate the data.

The resulting data set is flexible enough to be a gateway to a number of tools and ideas common to introductory (and some more advanced) courses. Students can be led to progressively deeper explorations as they advance in their knowledge and experience. Data production, sampling and bias, simulated sampling distributions, one- and two-sample hypothesis testing for proportions

and means (including matched-pairs testing and McNemar's test), confidence intervals, ANOVA, and linear regression are some of the topics to which the data set lends itself.

In addition to the aforementioned versatility, here are four potential benefits for students.

First, as teachers, we know that good research begins with a question. Students will be able to connect the survey to the academic research, exploring the origin of the ideas, and learning what kinds of questions others have asked and what has been done to answer them. Students can propose better ways to explore past questions or develop new questions of their own.

Second, students will encounter data production concepts, including sampling, bias, and replication in a realistic setting. There will be opportunities both to critique the methodology, and to suggest (or make) modifications, extensions, or wholesale changes.

Third, approaching a simple question and its corollaries from many perspectives adds richness to the investigation, and is closer to what researchers experience in the field.

Finally, and not least important, because the data is about the students themselves, they find it interesting and engaging.

2. Judging sex from handwriting

It seems obvious that when readers encounter a handwritten document they will tend to make judgments about its author, valid or not, based on the way the document is written. For example, might readers form conclusions about the sex of the writer based on the handwriting style? If so, might this influence the way they read the document? On the other hand, is it really possible to infer the sex of a writer from a sample of handwriting?

Attempting to identify personal traits (like gender) from handwriting is the aim of *graphology* ([Webster 2013](#)). Graphology is often classified as a *pseudoscience* in the following sense: though it is often treated as science, there is little to no evidence that personality or character traits can be discerned from a person's handwriting ([Beyerstein 1992](#)). This may come as a surprise to those who know of mystery novels, movies, and television shows that depict "experts" using handwriting to identify, for example, schizophrenia or homicidal tendencies. (Graphology is not to be confused with valid forensic uses of handwriting, like forgery detection or writing under duress.)

This brief discussion motivates the more direct initial questions we pose to students. "Do you ever make assumptions about the gender of a writer from a handwriting sample? Do you think you can, at least some of the time, determine gender from such a sample?" Students tend to respond with interest, and there are strong opinions both yes and no. In addition, most students would like to find out if they have some hidden skill.

A useful article by [Hartley \(1991\)](#) provides a summary of more than a dozen studies of handwriting recognition, conducted under a variety of conditions. These studies have reported average success rates for identifying gender from handwriting specimens of between 57% and

78%. A recent poll conducted by the consumer research company Survata ([Mathew 2013](#)) found an average of 54% correct identifications for 3100 respondents examining five handwriting specimens each. Of course, success rates will vary depending on the writing specimens chosen, and the sample of individuals surveyed. This obvious fact can trigger fruitful discussions about sampling techniques and bias.

3. Our survey

Our electronic survey contains 25 handwriting specimens collected from past undergraduate students aged 18 to 24. These students gave their permission, and anonymity has been ensured. There are 12 specimens created by male authors and 13 by females. Each is in the form of a mailing address label containing the same fictitious name and address. (See [Figure 1](#) for an example.) The specimens are displayed in random order, each followed by an identification question. (To avoid confusion with the dual use of the word “sample,” we will continue to refer to each sample of handwriting as a handwriting “specimen.”)

Figure 1. Survey question example.

E. Peterson
1034 North Sutton Street
Littlefield, VT 25796

This sample was written by a ...

Select one:

a. Male

b. Female

The address label format was inspired by three of the studies summarized in the article by [Hartley \(1991\)](#). An advantage is that no meaning is attached to the content that might influence the respondent. The format also helped blind the student writers to the purpose of the specimen collection (though they were told afterwards).

We administer the survey to each respondent (students in introductory statistics classes) twice, on different class days. We allow the respondents to believe something went wrong with the first administration and so they need to take it again. Since the survey takes an average of three to four minutes to complete, this is neither time-consuming nor cost-prohibitive, and enhances our ability to ask interesting questions, as we will demonstrate soon.

As we indicated in the Introduction, we prefer to administer the survey electronically. In the past, we have used an online survey product (surveymonkey.com), *StatCrunch*, or a course management system (Moodle). This semester, we intend to use a Google Docs form.

Using a course management system or a Google Docs form allows us to give the survey as a “quiz.” This has the labor-saving advantage of automatically transforming raw responses (male/female) to “scored” responses (correct/incorrect identification), thus enabling the immediate feedback students desire.

The resulting data is transferred into a spreadsheet for cleaning and organizing, and then into statistical software. We use *StatCrunch*, but many of the initial calculations can be done using the spreadsheet. These choices make the data set easily accessible to all the students, since access to MS *Excel* and *StatCrunch* are required for our course. At last count, our data set had 195 completed cases from students in past introductory classes. There are also eight partial cases from students who were only able to take the survey once.

The entire survey can be viewed at the URL in the footnote below¹, or by contacting the author if the reader would like a copy in another format. The raw and summary data and codebook are available at the JSE website at

<http://www.amstat.org/publications/jse/v23n1/bradley/HandwritingSummaryData.csv>;
<http://www.amstat.org/publications/jse/v23n1/bradley/HandwritingSurveyData.csv>, and
http://www.amstat.org/publications/jse/v23n1/bradley/Handwriting_and_Gender_documentation.docx.

4. Data summary

We use data gathered from the first three individual respondents to illustrate the variables of interest. These are shown in [Table 1](#). For example, individual respondent #1 was a female who made 72% correct identifications on Survey 1 (18 out of 25) and 68% on Survey 2 (17 of 25). The Female ID variable shows that when trying to classify female specimens on Survey 1, she was correct 75% of the time (9 of 12) and was correct 69.2% of the time (9 of 13) when identifying male specimens on Survey 1. We will explain the Both Correct variable later in the article.

Table 1. Summary results for three individuals.

Individual	Gender	Survey 1	Survey 2	Female ID	Male ID	Both Correct
1	Female	72	68	75.0	69.2	68
2	Male	56	68	41.7	69.2	48
3	Female	68	48	75.0	61.5	36

Of the 203 respondents, 112 are female (55.2%) and 91 are male (44.8%). Summary statistics for individual scores appear in [Table 2](#).

¹https://docs.google.com/forms/d/1sO6vlsozsORbqaCTsA7Ta0qZL7_6_MCEPJ7tYeKYyvI/viewform?usp=send_form

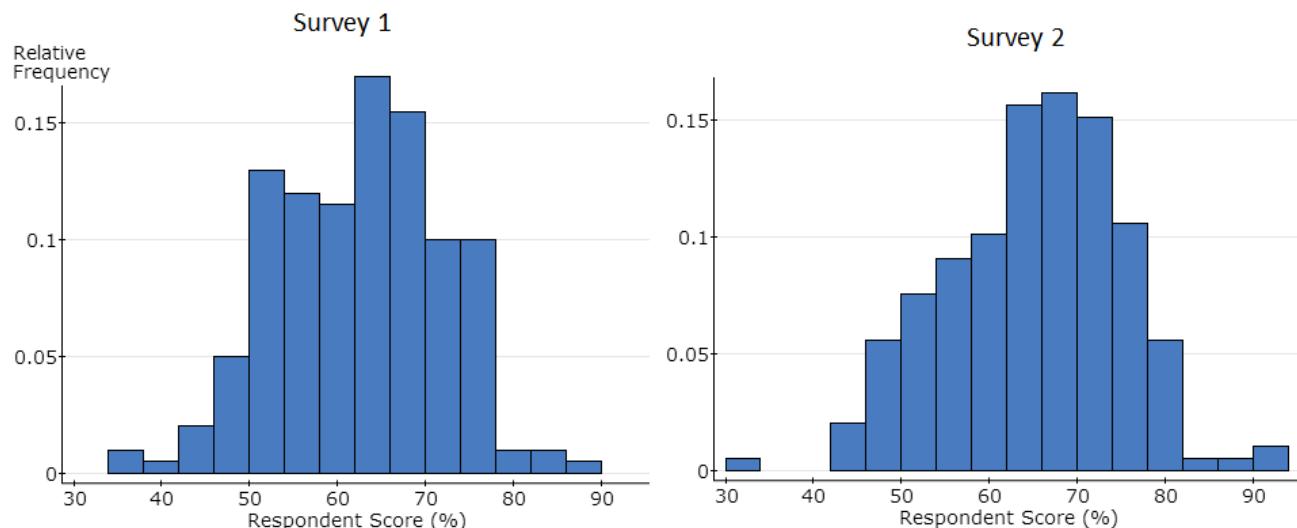
Table 2. Summary statistics.

	n	Mean	Std. dev.	Q1	Median	Q3	IQR
Survey 1	200	62.5	9.53	56.0	64.0	68.0	12
Survey 2	198	65.0	9.99	60.0	64.0	72.0	12
Female ID	203	63.9	15.99	58.3	66.7	75.0	16.7
Male ID	203	61.1	13.26	53.8	61.5	69.2	15.4
Both Correct	195	51.2	12.12	44.0	52.0	60.0	16

No respondent has yet achieved a perfect score. The minimum score is 8 out of 25 (32%) and the maximum is 23 out of 25 (92%), the latter being achieved only twice in nearly 400 attempts.

Helpful Hint: The relationship between the binary data collected and the quantitative scores for each respondent might at first be confusing to students. It might help to clarify that binary data are gathered from individual surveys. For each respondent, these data are translated into a percentage of correct responses. The means in [Table 2](#) are averages of these percentages.

Histograms for scores from Survey 1 and Survey 2 are shown in [Figure 2](#)². There are obvious differences in the distributions, even though each contained the same handwriting specimens.

Figure 2. Histograms of scores on the two identical (except for question order) surveys.

Helpful Hint: Ask students to compare the two histograms. “What are the similarities? What do you think accounts for the differences you see?”

² All the calculations and graphics were produced using *StatCrunch*, unless otherwise noted.

5. Individual students: Can you do better than a coin toss?

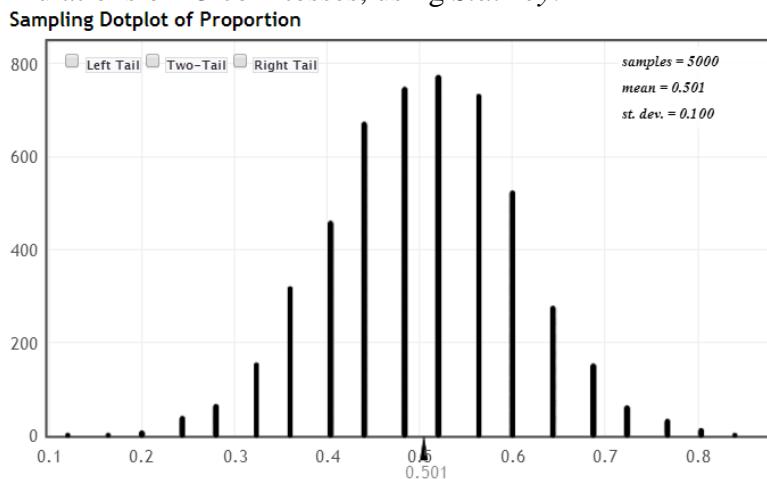
As we wrote earlier, the first question we asked students is the “hook”:

Can you tell the gender of a writer from a sample of his or her handwriting?

The act of taking the survey helps students to think more clearly about the question. They begin to realize there is some noise interfering with the signal of their ability to identify the genders of the writers. They become aware that some specimens are more difficult to pin down than others, and there is a certain amount of uncertainty involved. They have already divested themselves of the notion that they can identify handwriting in all cases, and perhaps even very well in general.

With this in mind, we begin by reframing the original question as, “Can you identify gender better than someone could by tossing a coin?” This bit of motivation sets the stage for the introduction of sampling distributions. We can use software to compute the percentage of correct identifications for thousands of simulated series of 25 coin tosses. In [Figure 3](#) we display the results of 5000 simulated sample proportions using *StatKey*³. We set the proportion for the simulations to 0.5 since a fair coin would be correct half the time in the long run.

Figure 3. Simulated sampling distribution of proportions of correct identifications in 5000 simulations of 25 coin tosses, using *StatKey*.



The context of computer simulation makes it easier for students to understand what information a simulated sampling distribution conveys: the results of many simulations of sets of 25 coin tosses to make guesses on the 25-question survey. The introductory texts by [Tintle et al \(2015\)](#), [Lock, Lock, Morgan, Lock, and Lock \(2013\)](#), and [Diez, Barr, and Cetinkaya-Rundel \(2012\)](#) all contain excellent expositions and suggestions for introducing many ideas via simulation-based sampling distributions.

³ *StatKey* is a collection of web-based statistics apps written to accompany *Statistics: Unlocking the Power of Data* by Lock, Lock, Lock, Lock, and Lock. The sampling distribution applet is available online at http://lock5stat.com/statkey/sampling_1_cat/sampling_1_cat.html.

Helpful Hint: Ask students to compare the histograms in [Figure 2](#) to the sampling distribution in [Figure 3](#). “What differences do you notice? What do you think accounts for these differences?”

Students readily observe the histograms in [Figure 2](#) have different centers than this simulated sampling distribution, and so they notice that respondents in general seem to have at least a little more ability than the computer does when it is “guessing.” Based on the simulated sampling distribution, it is not difficult to convince students they will have to score at least $17/25 = 68\%$ to be as good as the top 5% of all “lucky coin toss sequences,” and score at least $18/25 = 72\%$ to be part of the top 2.5%. This allows us to introduce informally (and not yet by name) the notion of a P-value as the probability a computer-simulated series of coin tosses does at least as well as the student did. It is then straightforward to follow up with more precise definitions within the context of hypothesis testing.

Such an idea can be introduced very early in a course. There is little essential background necessary to get students thinking about these ideas in the context of a binary variable. Once students have learned about sampling and bias, and the idea that probabilities can be estimated quite well via simulations, they are ready to approach sampling distributions. This is supported by the recent trend toward more randomization in the curriculum. See, for example, the article by [Tintle, VanderStoep, Holmes, Quisenberry, and Swanson \(2011\)](#).

If a student learns she performed better than coin tosses are expected to perform, then she might want to know how she compares to other humans. Was her performance better than average? Again, knowing there is some luck involved, we discuss with students that a score that might appear to be above average might just be the result of luck. Again, a simulated sampling distribution can help. Using the overall average so far (64%) as our best guess for the population’s mean, we use *StatKey* to create a new simulated sampling distribution where the overall proportion of correct identifications is $p = 0.64$. A simulated sampling distribution similar to [Figure 3](#) (but with $p = 0.64$) demonstrates the need to score $21/25 = 84\%$ or higher to be reasonably sure a student really does have *above average* talent and didn’t just get lucky. It would be even more convincing if such a student could *replicate* her success. That is one reason we give the survey twice, and we will discuss this in the next section.

Helpful Hint: Some students will recognize that the cutoffs for significance can vary with the randomization, especially for such a small sample size. For example, simulations here sometimes render the cutoff 84%, and other times 80%. This can lead to a discussion about how many simulations are needed to form an accurate sampling distribution. For students at an introductory level, we recommend at least 1000, and we tell them that there will be borderline cases that will force us to make difficult decisions. More advanced students are encouraged to explore the reciprocal relationship between the error and the square root of the number of simulations.

6. Questions to explore

Here we demonstrate the versatility of the data set. As various course topics arise, there are a number of questions students can explore. It is not our objective to perform all of the tests or

analyses our questions raise, but rather to provide a flavor of what is possible. Also, although we present the data we have collected in this article, we believe students are much more interested in their own data. For that reason, students explore some of these ideas using only the data collected by their particular class.

6.1 What is the population mean?

As stated earlier, Hartley (1991) provides a summary of more than a dozen studies, conducted under a variety of conditions, which report average success rates between 57% and 78%. The success rate is bound to vary depending on the specimens of writing chosen. So, how well do *our* respondents perform, on average? Using our data, *StatCrunch* gives the confidence intervals in [Table 3](#). If desired, comparisons could be explored easily via one-sample t-tests instead.

Table 3. *StatCrunch* confidence interval for Survey 1 and Survey 2 averages.

95% confidence interval results:

μ : Mean of variable

Variable	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
Survey 1	62.5	0.674	199	61.2	63.8
Survey 2	65.0	0.710	197	63.6	66.4

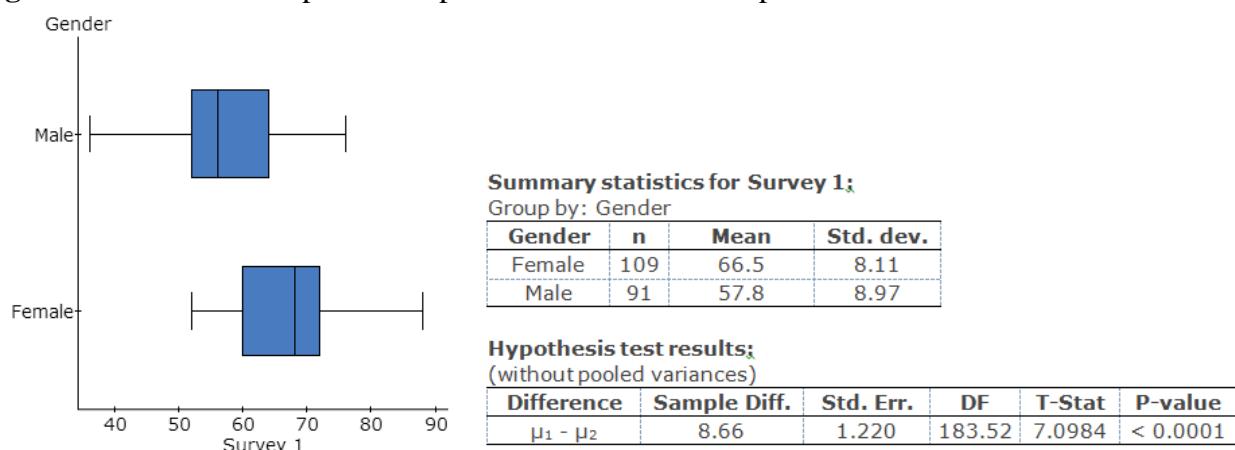
A discussion of population and sample is well-received at this point. What group is represented by the sample of respondents? How representative of the general population of writers is our particular collection of handwriting specimens? What improvements to either sampling methodology can be suggested?

6.2 How well did an individual student do?

Once one-sample significance tests have become part of the students' repertoire, testing the hypotheses $H_0: p = 0.50$ (to compare to computer-simulated "coin tossing") or $H_0: p = 0.64$ (to compare to the overall sample's proportion) can shed more light on how individual students performed on either survey. If confidence intervals have been covered, each student can construct one to learn about his or her "true score," accounting for at least some of the luck involved.

6.3 Do men's and women's scores differ significantly?

A two-sample t-test using the Survey 1 data suggests women are somewhat better than men at making identifications ($P < 0.0001$). See the *StatCrunch* output in [Figure 4](#).

Figure 4. *StatCrunch* output to compare male and female respondents.

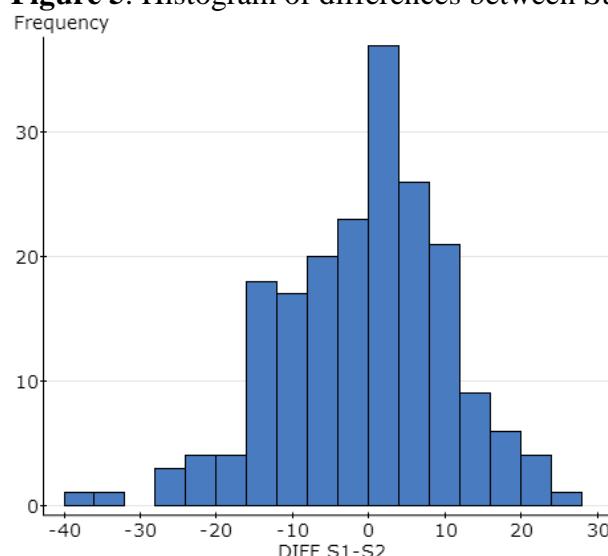
6.4 Is there a difference in scores between the two surveys?

[Table 2](#) reveals there is an approximate increase of 2.5 percentage points in scores between Survey 1 and Survey 2. Is this a significant rise? Many students want to perform a two-sample t-test to explore this, but the two samples are not independent of each other. This example helps students learn about the applicability of a matched-pairs (also called “repeated measures”) test. They can test the null hypothesis that the mean of the differences is zero. See [Table 4](#) and [Figure 5](#) below, where the new variable DIFF S1 – S2 refers to an individual’s difference in scores on the two surveys.

Table 4. *StatCrunch* results for significance test of $\mu_D = 0$.

Hypothesis test results:

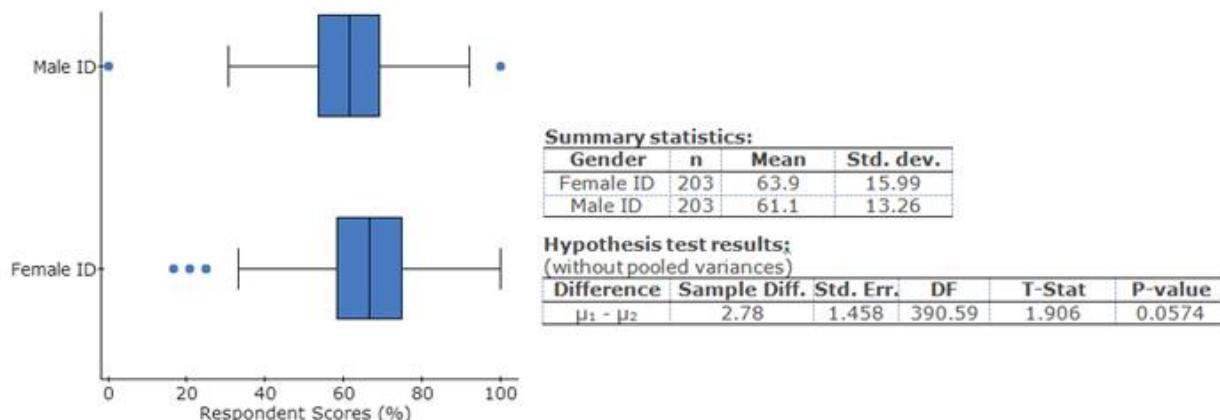
Variable	Sample Mean	Std. Err.	DF	T-Stat	P-value
DIFF S1-S2	-2.56	0.786	194	-3.264	0.0013

Figure 5. Histogram of differences between Survey 1 and Survey 2 scores.

6.5 Do respondents perform differently when identifying male vs. female handwriting specimens?

In [Figure 6](#) we display the *StatCrunch* output to compare the performances of respondents when they are judging only female specimens (FemaleID) vs. only male specimens (MaleID). Though there is some evidence that respondents are better at identifying female handwriting (the *StatCrunch* two-sample comparison of means in [Figure 6](#) reveals that $P = 0.0574$), it is not compelling (or statistically significant at the 0.05 level). Alternatively, students could use confidence intervals to explore this question.

Figure 6. *StatCrunch* output comparing responses to female vs. male specimens.



6.6 Do respondents perform differently when examining specimens of their own, or the opposite gender?

The previous question about whether students are better at identifying one specific gender's specimen than the other's leads naturally to the question of whether respondents of one gender are better at recognizing handwriting specimens of their own or the opposite gender. [Table 5](#) suggests that when attempting to identify female specimens, males and females have markedly different success rates. We leave a specific (ANOVA) test for the interested reader.

Table 5. *StatCrunch* output comparing mean scores (and standard deviations) for respondent gender vs. specimen gender.

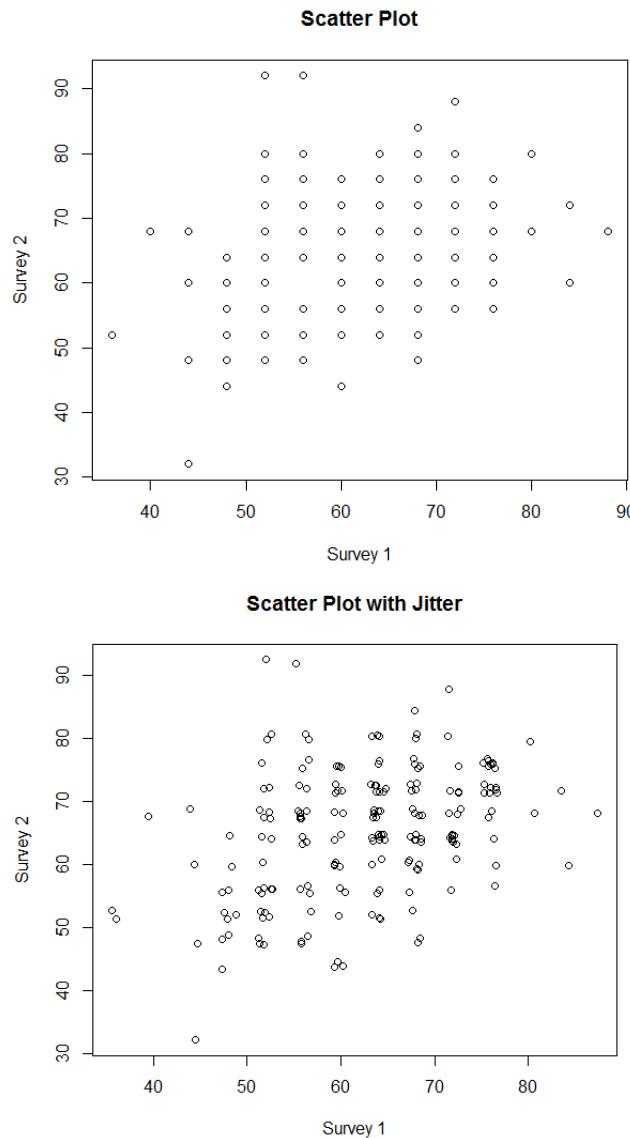
Respondent	Handwriting Specimen	
	Male	Female
Male	60.1 (15.71)	55.6 (16.76)
Female	61.9 (10.88)	70.6 (11.66)

6.7 What is the effect of randomness on individual scores?

We mentioned earlier the differences in the histograms for the two survey versions, despite the specimens being the same (but presented in different orders). For many students, this confirms

their suspicion that there is some noise interfering with the signal of an individual's level of skill in identifying genders of writers. (Students may also have noticed the difference in the confidence intervals for Survey 1 and Survey 2, as shown in [Table 3](#).)

Figure 7. Scatterplots with and without jitter, showing individual scores on the two surveys.



Here, an exploration of a scatterplot can be fruitful. Since so many data pairs are repeated, we show in [Figure 7](#) scatterplots created in R both with and without jitter. (The jitter feature is not available in *StatCrunch*.) The scatterplots inspire more questions students can explore:

- What does the amount of scatter reveal? The correlation is $r = 0.376$. Though there is a clear positive trend, there is also a substantial amount of variation in individual retakes of the survey, and the relationship is not as strong as most students suspect it should be. The variation is also evidenced by the standard deviation of the differences in survey scores ($\hat{\sigma}_D \approx 11$).

- If respondents score the same on both surveys, how is this revealed in a scatterplot? Does this happen often? The $y = x$ line is relevant here. Of the 195 respondents who completed the survey twice, 37 scored the same both times (about 19%).
- How many of the scores are above or below the $y = x$ line? There are 91 and 67, respectively. What does this reveal about the differences in the two survey attempts? Students can perform a one-sample test of proportions to determine if there is a difference in the proportion of respondents who scored better on Survey 2 than for those who scored better on Survey 1. Of these 158 respondents, $\hat{p} = 91/158 = 0.576$ scored higher on Survey 2. This is not significantly more than half (two-sided binomial $P = 0.0669$).
- How many respondents scored 80 or above on Survey 1? On Survey 2? On both surveys? (The answers here are 5, 15, and 1.) What do you think this means? The data help to illustrate regression to the mean – it is more difficult to “get lucky” twice. Respondents who scored unusually high or unusually low on one survey tend to score much closer to the mean on their other attempt.

The discovery of so much variation in the scatterplot led one student to suggest a more stringent scoring method: an individual is credited with a correct identification only if he responded correctly on both surveys. (These are the scores in [Tables 1](#) and [2](#) labeled “Both Correct.”) With this replacement, the mean score drops to 51.2%. With this added stipulation, only a handful of students, 20 out of 195, performed significantly better than a coin-tossing simulation would be expected to (that is, 68% or higher)! And this still does not rule out luck entirely.

At this point, some students find it interesting to revisit the questions listed above using the “Both Correct” scores.

6.8 Is there a difference in the proportions of changed responses?

We can extend the above investigation by exploring individual responses which differed on the two surveys. For example, consider the two-way table below. Ignoring the $496 + 253 = 749$ responses that were the same on both surveys, we can explore a potential trend among the changed responses, determining if individuals were more likely to change from correct to incorrect, or vice versa.

Table 6. Tallies of incorrect and correct responses cross-checked by survey, for 42 respondents during one semester.

		Survey 1	
		Correct	Incorrect
Survey 2	Correct	496	153
	Incorrect	148	253

Respondents changed their responses from an incorrect to a correct response 153 times, and from a correct to an incorrect response 148 times (the total number of changes is 301). We can use McNemar’s Test to determine whether there is a significant difference in these two proportions (see, for example, [Utts and Heckard \(2012\)](#)). In this case an exact binomial test will provide equivalent results: there is not a significant difference between the two proportions ($P = 0.730$).

When students respond differently, there is no evidence they are more likely to be correct, or incorrect, more often the second time around.

This kind of exploration can be conducted in two other contexts here as well. It can be used to focus on an individual handwriting specimen. When individuals change their responses, are they more likely to get it right (or wrong) the second time around? Or it can be used to learn about a single respondent. Is there a significant difference in the way a single subject responded on the two surveys across the 25 questions?

7. Adaptations and extensions

There are a variety of avenues we might encourage students to explore. Some require collecting more data, or altering the survey to ask for other types of information.

- Focusing on the handwriting specimens, examine the percentage of correct identifications of each handwriting specimen. Two of our specimens are identified correctly more than 90% of the time. Two others are correctly identified less than 5% of the time. What can we learn from the distributions of these scores? What can be learned about whether a particular specimen “looks like” a female specimen or a male specimen to most respondents?
- The survey can be altered so that more response options are available, possibly to include a scale like:
 - ◆ I’m very sure this was written by a **male**.
 - ◆ I’m somewhat sure this was written by a **male**.
 - ◆ I can’t tell.
 - ◆ I’m somewhat sure this was written by a **female**.
 - ◆ I’m very sure this was written by a **female**.

Assigning these responses to a Likert scale could lead to a “gender perception” score assigned to each specimen, and these values could be investigated further.

- Students can alter the collection of handwriting specimens to explore the effect these have on responses. There is bound to be some dependence on which handwriting specimens are used. How much will the results vary when different writers are used? And does the content of the specimen matter? That is, might certain letters or words serve as triggers for respondents?
- Students could create an experiment where they try to fool each other, writing as if they were of the opposite gender to their own (whatever they might think that means).
- Other demographic variables could be included in the survey to explore relationships between a new variable and the scores of respondents (handedness, age, major, ...).
- The specimens themselves could come from different age groups: middle schoolers, senior citizens, etc.

8. Access to surveys

Readers can contact the author for the location of the surveys in *StatCrunch* or surveymonkey.com, and for access to the data once their students have completed them. Alternatively, instructors can create their own surveys using their school's learning management system, or through an online survey provider like SurveyMonkey. We believe students are more interested in analyzing personal data.

9. Conclusion

Though at first just a fun idea for students to play around with, the combined data sets proved to be unexpectedly rich, especially when we thought of asking students to try to replicate their first handwriting identification score. We are aware there is a danger of over-exposure; students can lose interest if a data set becomes the sole focus of the course. But many students find these questions interesting enough that they want to explore them, either as part of class or as guided independent projects.

It is worth noting that one can obtain similarly rich data sets without making handwriting recognition the focus. One could replace handwriting recognition with any skill assessment that involves some aspect of randomness to conduct analyses similar to what we have suggested here. All that is needed is a test students can take more than once.

Acknowledgements

We are grateful to students who have “donated” their handwriting specimens, and those who have taken the survey and worked with these ideas in our statistics classes. We are especially grateful to two future teachers, Matthew Naber and Stephanie Spillane, who offered advice and many good suggestions about how to improve this project for future generations of students.

Though this was deemed to be an “exempt” study for research purposes, we requested and received approval from Clarke University’s Institutional Review Board.

References

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, B., Utts, J., Velleman, P., and Witmer, J. (2012), “Guidelines for Assessment and Instruction in Statistics Education: College Report” [online]. American Statistical Association. Available at http://www.amstat.org/education/gaise/GaiseCollege_Full.pdf.

Bailey, B., Spence, D. J., and Sinn, R. (2013), “Implementation of Discovery Projects in Statistics,” *Journal of Statistics Education* [online], 21(3). Available at <http://www.amstat.org/publications/jse/v21n3/bailey.pdf>.

Beyerstein, B. (1992), *The Write Stuff: Evaluations of Graphology, the Study of Handwriting Analysis*, Buffalo, N.Y.: Prometheus Books.

Diez, D. M., Barr, C.D., and Cetinkaya-Rundel, M. (2012), *OpenIntro Statistics* (2nd ed.) [online]. Available at <https://www.openintro.org/stat/textbook.php>.

"Graphology." *Merriam-Webster Dictionary* (2013), [online]. Available at <http://www.merriam-webster.com/dictionary/graphology>.

Hartley, J. (1991), "Sex differences in handwriting: A comment on Spear," *British Educational Research Journal*, 17(2), pp. 141-146.

Lock, R. H., Lock, P. H., Morgan, K. L., Lock, E. F., and Lock, D. F. (2013), *Statistics: Unlocking the Power of Data*. Hoboken, NJ: Wiley.

Mathew, T. (2013), "Identifying gender by handwriting: you're probably not as good at it as you think," *Survata Blog*, December 6, 2013, [online]. Available at <http://survata.com/blog/identifying-gender-by-handwriting-youre-probably-not-as-good-at-it-as-you-think/>.

Tintle, N., VanderStoep, J., Holmes, V., Quisenberry, B., and Swanson, T. (2011), "Development and assessment of a preliminary randomization-based introductory statistics curriculum," *Journal of Statistics Education*, 19(1). Available at <http://www.amstat.org/publications/jse/v19n1/tintle.pdf>.

Tintle, N., Chance, B. L., Cobb, G.W., Rossman, A. J., Roy, S., Swanson, T., and VanderStoep, J. (2015), *Introduction to Statistical Investigations*. Hoboken, NJ: Wiley.

Utts, J., and Heckard, R. (2012), *Mind on Statistics*. Boston, MA: Brooks/Cole Cengage Learning, p. 601.

Sean Bradley
Clarke University
1550 Clarke Drive
Dubuque, IA 52001 USA
sean.bradley@clarke.edu
