

Teaching Quantitative Methods Vignettes

In-class activity comparing standard errors as a function of sample size using SPSS

Denis Cousineau ^a

^aUniversité d'Ottawa



Abstract ■ This vignette presents an in-class demonstration to introduce the notion of standard error. The students must be able to compute means using SPSS. They must also be able to eye-ball standard deviation from a list of numbers. In doing the activity, they will be exposed to sampling. The demonstration uses a random number generator so that each student faces a different dataset simulating IQs from a single group of participants. Factors affecting standard errors (sample size and population variance) are explored. Finally, the actual formula for standard error is then compared to the roughly estimated standard deviation of the sample means..

Keywords ■ Quantitative vignette; statistics; standard error. **Tools** ■ SPSS.

denis.cousineau@uottawa.ca

[10.20982/tqmp.16.2.v009](https://doi.org/10.20982/tqmp.16.2.v009)

Acting Editor ■ Sébastien Bélard (Université de Montréal)

Reviewers ■ No reviewers.

Concept to be presented

The purpose of this vignette is to introduce an activity that can be performed in class under the supervision of the instructor. The purpose is to make student realize that larger samples are more reliable than small samples, which is exemplified by the standard error. The standard error of the mean in a single-group design is quantified using a well-known formula

$$SE_{mean} = \frac{SD}{\sqrt{n}} \quad (1)$$

in which SD is the sample's standard deviation and n is the sample size. The purpose of this activity is not so much one in which the students must apply the above formula. It is rather one in which the students realizes that estimation error is smaller when samples are large, and vice versa. To that end, we propose to use sampling from a simulated population. Using a computer, we ask students to generate (simulated) samples and ask them to compute the means. Then the means are reported to the instructor and written down on the board. By comparing means from two sets of simulation (large n simulations and small n simulations),

the students are invited to realize how precise the former estimates are relative to the second.

Activity

In the class, have students be prepared with SPSS open. Prior to the class, the free and open-access extension GRD (*Generate Random Data* Harding & Cousineau, 2014, for version 1; or Harding & Cousineau, 2015, for version 2) must be installed (version 2 has a graphical user interface; we present step-by-step instructions for either the syntax mode –both versions– or the interface mode –only version 2).

Start the activity by discussing the notion of precision. Running a study at estimating a population mean characteristic. For example, let them know that you would be interested in assessing the IQ of a group or person met on the street. As they certainly know, the average IQ is – by the way the questionnaire is built– 100 and people tend to differ from the mean by about 15 points (a rough intuition regarding standard deviation). At this point, invite the student to discuss in what circumstances they should get a better estimate from a sample of people. They should

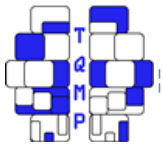


Table 1 ■ Example of means found by students who generated small (left column) or large (right column) samples of IQs. Variability is roughly 5 times larger on the left than on the right.

Small sample's means	Large sample's means
98.6	100.1
106.2	100.5
97.4	100.7
91.1	99.6
100.4	98.2
92.8	101.5
93.7	98.2
98.2	100.2
108.2	98.0
97.8	101.6

rapidly suggest that a large sample is better than a small sample.

At this point divide the class in two. The left part will be invited to collect IQs from 9 persons whereas the right part of the class will collect IQs from 225 persons. Each student will be in charge of collecting one sample of its assigned size.

To that end, pretend that you have a module connected with Google which aims at recognizing the IQ of its users and which has access to millions of IQs. This module is called the Google's Recognition of Gifted persons (GRD; if you have used GRD earlier in the semester; they will understand that it isn't true; however, if this is the first time, be wary that some students will take your word literally).

With this module you can ask to get the IQ of a sample of Google users of the size you desire. They just have to type in the syntax given in Listing 1 or use the graphical user interface given in Figure 1. The left-side of the class keeps the number "9" whereas the other part of the class replaces "9" with the sample size "225".

Upon running the above instructions, each student should get a different sample. Let have them realize that by looking at their neighbors' data shown in the data editor. They may also visualize the frequency distribution plot to better appraise the differences between samples if they are familiar and know how to make this plot.

Now the focus of the exercise is on the sample means. Let the student compute the mean (using either the syntax MEANS IQ. or the menu Analyze: Descriptive Statistics: Descriptives...). As the results are obtained, report them on the board in front of the class, keeping the two halves separated (small samples vs. large samples). Table 1 shows one possible result of this exercise.

As seen, the numbers on the left columns are considerably more variable than the number on the right. On the right, deviation to the true population mean is often visible only on the first decimal whereas on the left, deviations are

patent.

Invite the student to eye-ball the magnitude of the typical deviation to 100 on the left column. They should suggest a typical deviation of about 5 points (some are larger than 5 points away, some are smaller). On the right, they might suggest roughly one point of deviation more or less.

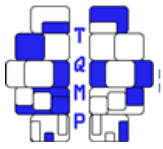
Start a discussion on the meaning of these two numbers (5 and 1). Have them realize that this is a measure of imprecision (lack of precision) and that a large number is an undesirable result.

Finally, use Eq. 1 above to show that you could predict these results just by knowing the sample size of one sample and its standard deviation. By using 15 as a measure of the population standard deviation, and either 9 (small samples) or 225 (large samples), you find 1.0 and 5.0, in good agreement with what the results across multiple samples indicates.

The strength of Eq. 1 is that a single sample is sufficient to infer what is found by looking across multiple samples. Thus, Eq. 1 is a sort of magical formula, able to anticipate the amplitude of imprecision if multiple samples had been collected, even though in practical applications, we never collect multiple samples.

If time permits, you can have the student explore an even more extreme sample size (a square number such as 400, 900 or –why not!–10,000, to facilitate computation of Eq. 1). They can do this additional exercise on their own; they should all observe that precision is even better.

If your students are well familiarized with SPSS, you can expand the syntax of Listing 1 such that each student generates multiple samples, called Replications. See Listing 2 for a modification that permits the creation of 100 samples, each having 9 scores (that represents 900 lines in the data editor; SPSS can generate up to a hundred replications containing hundreds of thousands of random data in less than a few seconds).



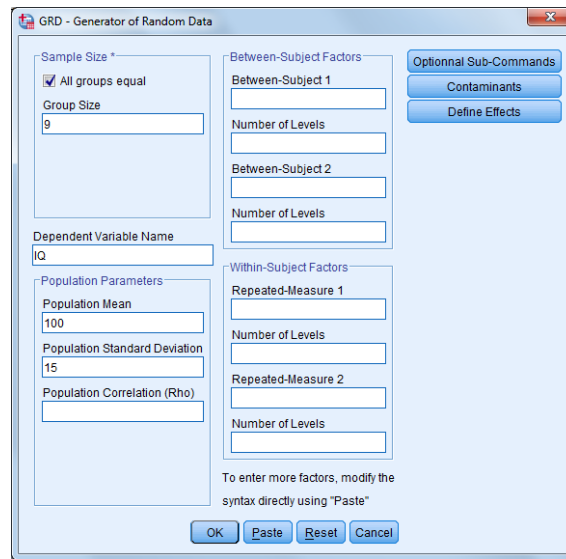
Listing 1 ■ A syntax for SPSS to generate 9 random scores simulating IQs. It is based on Harding and Cousineau (2015) extension to SPSS but works equally well with version 2 (Harding & Cousineau, 2016).

GRD

```

/SUBJECTSPERGROUP equal = 9
/RENAME DV = IQ
/OVERALL MEAN = 100 STDDEV = 15.
    
```

Figure 1 ■ A graphical user interface for SPSS to generate 9 random scores simulating IQs. It is available only with version 2 of GRD extension to SPSS (Harding & Cousineau, 2016).



Key concepts to remember

Standard error is a measure of precision. More precisely, it is a measure of imprecision, with large standard error representing potentially imprecise samples. In real life situations, it is not possible to know whether the sample mean is actually precise or not. However, a large standard error indicates that it is likely not precise. Standard error can be obtained from a single sample but measures how much a large number of samples would fluctuate. It is therefore a shortcut.

This formula applies in cases where sampling is fully random (i. e., anyone in the population has an equal chance of being selected) and also when the population is large of infinite. When these prerequisite are not met, adjustments to this base equation are possible (Cousineau, 2017).

Follow-up suggestions

A possible follow-up to this exercise would be to fix the sample size (to say 25) and invite the student to find an al-

ternative way to improve precision of the mean estimate.

If the students are comfortable, a formal mathematical demonstration of the formula for the standard error of the mean can be provided. Such a demonstration is college-level and found in many textbooks (e. g., Cousineau, 2010; Hays, 1973).

Authors' note

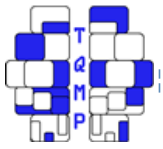
The author wishes to thank Sébastien Béland, Michael Cantinotti, Marc-André Goulet, Bradley Harding, Yves Lacouture and Samuelle Rochette for useful discussions.

References

Cousineau, D. (2010). *Panorama des statistiques pour psychologues*. Bruxelles: Les éditions de Boeck Université.

Cousineau, D. (2017). Varieties of confidence intervals. *Advances in Cognitive Psychology*, 13, 140–155. doi:10.5709/acp-02140z

Harding, B., & Cousineau, D. (2014). Grd: An spss extension command for generating random data. *The Quantita-*



Listing 2 ■ A syntax for SPSS generating 100 samples of 9 scores simulating IQs and then computing the means within each replication. If you are using the graphical user interface of Figure 1, add to the box "Between-subject 1" the name "replication" and in the box "Number of levels" that follows, "100" to achieve the same result.

GRD

```
/BSFACTORS replication (100)
/SUBJECTSPERGROUP equal = 9
/RENAME DV = IQ
/OVERALL MEAN = 100 STDDEV = 15.
```

MEANS IQ by replication.

tive Methods for Psychology, 10, 80–94. doi:[10.20982/tqmp.10.2.p080](https://doi.org/10.20982/tqmp.10.2.p080)

Harding, B., & Cousineau, D. (2015). GRD 2.0: An extended SPSS extension command for generating random data. *The Quantitative Methods for Psychology, 11*, 127–138. doi:[10.20982/tqmp.11.3.p127](https://doi.org/10.20982/tqmp.11.3.p127)

Harding, B., & Cousineau, D. (2016). GSD: An SPSS extension command for sub-sampling and bootstrapping datasets. *The Quantitative Methods for Psychology, 12*, 145–153. doi:[10.20982/tqmp.12.2.p145](https://doi.org/10.20982/tqmp.12.2.p145)

Hays, W. L. (1973). *Statistics for the social sciences*. New York: Holt, Rinehart and Winston, inc.

Citation

Cousineau, D. (2020). In-class activity comparing standard errors as a function of sample size with SPSS. *The Quantitative Methods for Psychology, 16*(2), v4–v7. doi:[10.20982/tqmp.16.2.v009](https://doi.org/10.20982/tqmp.16.2.v009)

Copyright © 2020, Cousineau. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 26/03/2019 ~ Accepted: 09/02/2020

Extended activity metadata

<i>Concept illustrated</i>	Standard error of the mean	<i>Type of activity</i>	In-class with instructor
<i>Prerequisite</i>	Sampling	<i>Types of data</i>	Computer-generated
<i>Co-requisite</i>	Sample mean, variability	<i>Computation by</i>	SPSS
<i>Suitable class size</i>	Medium (25-55)	<i>Duration</i>	20 minutes