An informative, concise title

Author’s name and the date

Introduction

Start with a statement introducing the problem to be explored and motivating the experiment. Refer to past readings or modules in which we explored the topic whenever possible, summarizing their content. This section should answer the question, “Why do we care about this topic?” State the purpose/objectives of the lab exercise.

Instruments

State the name, brand, model, and manufacture year of instrument, if available. State the accuracy of the instrument as given by the manufacturer.

Methods

State the location, date, and start and end times of the data collection. Describe how the data were collected, being specific. Illustrate with a diagram, photograph, or video, if possible. Remember: readers should be able to reproduce the experiment based on your description.

Data

In this section, you can discuss the characteristics of the data, not your interpretation of them. (That comes in the next section.) If possible, include the actual data in this section. If including all the data points would make the report an unreasonable length, you may include the data in tabular, graphical, or descriptive statistical form instead. (Modern scientific papers may include large data sets and/or open source analysis software as an appendix or electronic supplement.)

Discussion

This section is for explaining your interpretation of the data (including reasons for excluding any alternative hypotheses), any assumptions, and any limitations of the results.

Conclusions

Briefly restate what was done and the principal findings. Tie this section back into the introduction, bringing the reader back to the main purpose of the experiment. How have your actions have resolved the knowledge gap you described in the introduction?

Acknowledgments

If you received assistance, support, or funding from anyone in the data collection, analysis, or preparation of the lab report, acknowledge them here. This is also the place where you list any software used, with version numbers.

References cited

This is the bibliography for the lab report. You must cite relevant literature when it is appropriate to do so. Such citations give credit to those who have done research before on your topic, and create an evidence trail that aids in reproducibility.

Cite *primary, peer-reviewed* sources whenever possible. Wikipedia may be where your research begins, but not where it ends.

Different scientific journals have different requirements for citation and bibliographic style. Use AMS-formatted references to literature or instrument manuals, using DOIs whenever possible. Citation management programs such as Zotero can be set up to produce properly formatted citations and bibliographies, and integrate nicely with Google Docs and Microsoft Office 365.

There should be no bibliographic entries for items that are not cited in the paper. Conversely, there should be no citations for items not listed in the bibliography.

Notes regarding figures, tables, and equations:

- Figure captions go below the figure; table captions go above the table.
- Figures, tables, and their captions should not be split across page breaks.
- All figures, tables, and equations should be referenced in the text, and numbered in the order in which they are presented. If you don't discuss it, then don't include it.
- If you annotate the figures, make sure the annotations don't cover up the data!

Other writing tips:

- **Remember that the purpose of the lab report (or any scientific paper) is reproducibility.** I.e., a reader should be able to reproduce your experiment reasonably well by reading your instructions. However, you should take care not to bore the reader with extraneous details that are not important to the experiment (e.g., the use of gloves while handling thermometers is important, but the color of the gloves is not). There is a delicate balance between *detail* and *concision*.
- The word "data" is plural. The word "datum" is singular. (The phrase "data set" is also singular.)
- "i.e." means, "that is." "E.g." means, "for example."
- The experiment and calculations should always be described in the past tense, because they were performed in the past relative to the time of writing.
- Integer numbers less than 10, or those that occur at the beginning of a sentence, should be spelled out (e.g., "five measurements" or "two hygrometers"). Numbers greater than 10, or those containing decimals, should be given in Arabic numerals (e.g., 13 thermometers, or 6.5 °C)
- AMS style is to use **SI units** and **UTC time**, but since we won't take data spanning multiple time zones in this course, it's okay to use local time (EDT) as long as you give a UTC equivalent on first use. If in doubt, hit up time.gov to get the current time offset for UTC.

"Tell 'em what you're gonna tell 'em, tell 'em, and then tell 'em what you told 'em." -- Lou Wicker

Reducing systematic error in liquid-in-glass temperature measurements

Robin L. Tanamachi, Ph.D.
EAPS 22700 Lab 1
5 September 2021

Introduction:

Atmospheric measurement errors can be separated into two types: systematic and random. As we learned in the MetEd Module "Foundations of Meteorological Instrumentation and Measurements" (UCAR/COMET 2017), a principal objective in any atmospheric measurement campaign is to minimize the effects of systematic errors (which are usually preventable) on measurements, so that only random errors (which cannot be prevented) manifest in the measurements.

In this laboratory activity, we attempted to accurately measure the temperature in our classroom using a set of 10 liquid-in-glass thermometers. The impact of systematic error on our measurement was quantified by rectifying systematic error midway through the exercise.

We tested the following hypotheses:

1. Consistent measurement technique will reduce the spread (standard deviation) of the temperature measurements.
2. Measurements taken before and after systematic error is eliminated will be different enough to be considered two different data sets.
3. Quality control will reduce the variance in the "before" data set, but not as much as reducing systematic error does.

Instruments:

We used a set of 10 Easy-Read brand liquid-in-glass thermometers containing dyed ethyl alcohol. According to the instrument manufacturer's Traceability Statement (Appendix A), the accuracy of these thermometers within the range 0 °C to 100 °C is half a gradation, or ±0.5 °C.

Methods:

Data collection commenced at 10:00 a.m. EDT (1400 UTC) on 25 August 2021, in Room 231 of Stanley Coulter Hall, and ended at 10:20 a.m. EDT. Following Purdue COVID protocols, students wore nitrile gloves to prevent cross-contamination of the thermometers as they were passed around. Participants were not given instructions for reading the thermometers before data collection began other than to circulate 10 tagged thermometers among themselves, read them by eye, and record the measurements.

Students were asked to enter four parameters into a collaborative spreadsheet: the measured temperature, their initials, and the thermometer number, before passing the thermometer to a nearby person. Timestamps were automatically appended to each entry.

After five minutes, the instructor stopped the data collection and gave students instructions for how to properly read the thermometers. These instructions included:

- Reading thermometers at eye level (thereby eliminating parallax error).
- Handling thermometers by the stems rather than the bulbs (thereby removing one source of exposure error).
- Keeping the thermometer bulbs away from participants' mouths to avoid heating by breath (another source of exposure error).
- Basing the temperature measurement on the bottom of the concave meniscus (thereby reducing illegitimate error).

The participants then circulated the thermometers for an additional five minutes and recorded their temperature measurements in the same spreadsheet. The data set was therefore easily separable into "before" and "after" subsets relative to when the instructions were given.

Three remote participants acted as quality control officers for the data set. They subjectively flagged spreadsheet rows as "suspect" (1) if their measurements appeared to be outliers, or if the entries were missing any of the four required fields. Entries not meeting these criteria were flagged as "good" (0).

Data:

The time series of air temperature data from classroom SC 231 are displayed in Fig. 1, and basic characteristics of the data are given in Table 1.

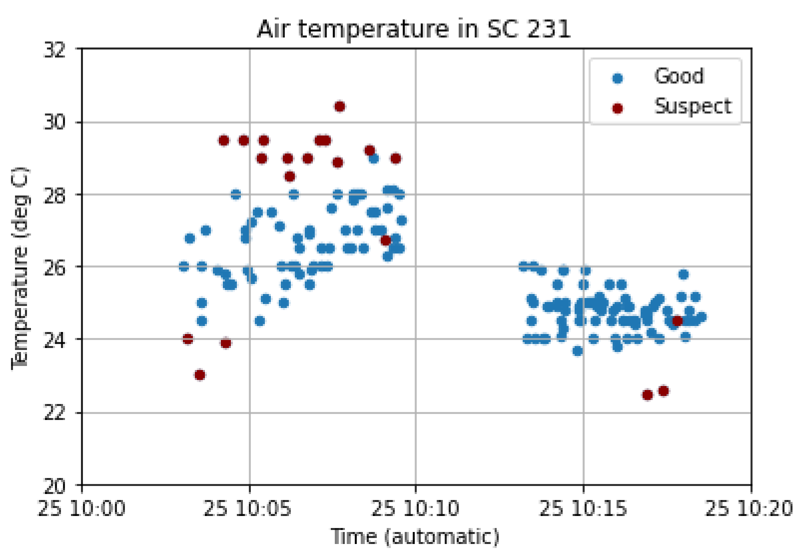


Fig. 1. Air temperature measurements (in °C, colored dots) as a function of time (EDT). Measurements flagged as "good" are plotted in blue; "suspect" measurements are plotted in red. Instructions were given at 10:10 a.m. EDT.

Table 1. Characteristics of the data set.

	Measurements collected	Mean measured temperature (°C)	Standard deviation (°C)	Suspect measurements (QC flag = 1)
Before instructions	84	27.0	1.5	18
After instructions	81	24.7	0.6	3

Discussion:

Hypothesis 1: Systematic error is reduced by consistent, correct measurement technique.

It is evident from Fig. 1 and Table 1 that the "before" and "after" data subsets have very different statistical characteristics, despite having comparable sizes (84 and 81 measurements, respectively; Table 1). First, the "before" subset had considerably more "suspect" data points (18, or 21%) than the "after" subset (3, or 4%). Second, the mean temperature measurement before instruction was more than 2.0 °C higher than that after instruction (Table 1). Third, the standard deviation of the temperature measurements before instruction (1.5 °C) was considerably higher than that after instruction (0.6 °C, Table 1), indicating that there was greater variability in those measurements before instruction. Notably, after instruction, the standard deviation of the temperature measurements was nearly equal to the manufacturers' stated accuracy for the thermometer (0.5 °C), lending credence to the idea that instruction minimized systematic error (and hence, that random error became the dominant source of error). <u>We therefore accept Hypothesis 1.</u>

Hypothesis 2: The "before" and "after" data sets are statistically distinct.

As a first check, we inspected the interquartile ranges of the two data subsets (Fig. 2). They do not overlap, indicating that the two subsets represent different data.

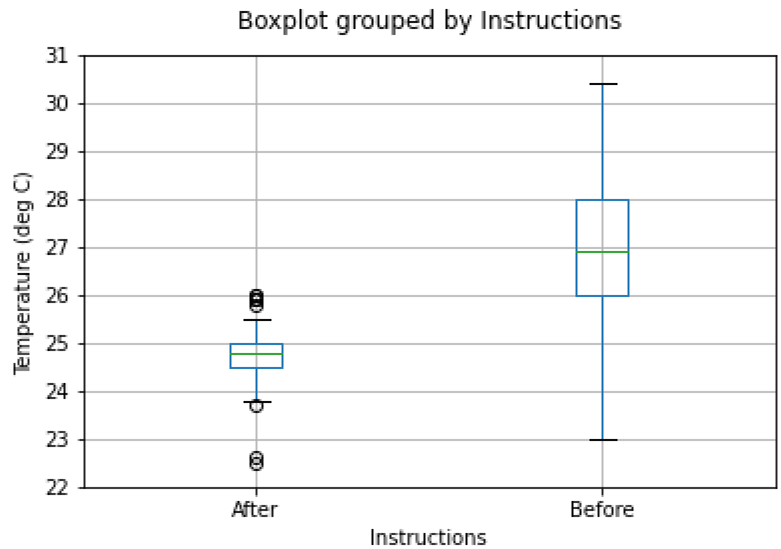


Fig. 2. Box-and-whisker plots of the temperature data (in °C) collected before and after 10:10 a.m. (when instructions were given).

Second, we used a two-sample Student's T-Test in the context of a repeated measures design (Kalpić et al. 2011). This situation satisfies the validity conditions for a two-sample Student's T-test, specifically:

1. The two data sets are independent of one another (i.e., measurements are not paired).
2. The variances of the two data sets can be *expected* to be equal (even though we found that they weren't when testing Hypothesis 1), because we used the same thermometers to take both subsets.

The null and alternative hypotheses were formulated as

$$H_0 : \mu_{Before} = \mu_{After} \tag{1}$$

and

$$H_a : \mu_{Before} \neq \mu_{After}, \tag{2}$$

respectively, where μ signifies the mean of a data set.

We applied the Student's T-test function built into the SciPy (Virtanen et al. 2020) package for Python. The resulting T-statistic was 12.9 and the p-value was on the order of 10^{-26}, indicating with high certainty that the null hypothesis H_0 should be rejected (threshold: $p < 0.05$). In other words, the odds of our having obtained the "before" and "after" data that we did by chance if their means truly were equal is vanishingly small. We therefore consider the alternative hypothesis H_a -- that the means of the two data sets are <u>not</u> equal -- to be 100% likely. <u>We therefore conclude that the "before" and "after" data sets are distinct, affirming Hypothesis 2.</u>

Hypothesis 3: Reduction in systematic error is more valuable than post-experiment quality control.

An ancillary experiment was conducted using the QC flags provided by the remote participants. As previously mentioned, far more measurements were flagged as “suspect” in the “before” data (18, or 21%) as in the “after” data (3, or 4%). To test Hypothesis 3, we removed all of the suspect data points from both data subsets and regenerated the interquartile ranges (Fig. 3) and statistics (Table 2).

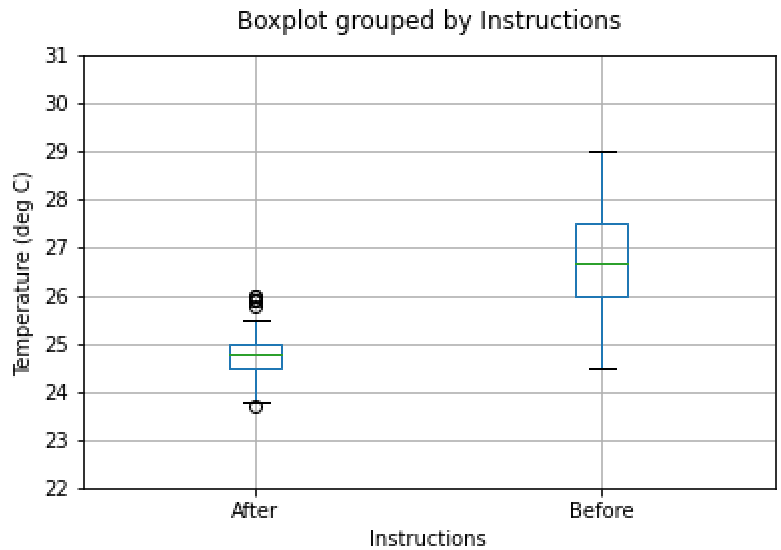


Fig. 3. Same as Fig. 2, but with suspect data removed.

Table 2. Characteristics of the data set following removal of suspect data.

QC flag = 0 only	Measurements collected	Mean measured temperature (°C)	Standard deviation (°C)
Before instructions	66	26.6	1.0
After instructions	78	24.7	0.5

Removing the suspect data decreased the standard deviation of both data sets (Table 2). In particular, the standard deviation of the “before” data decreased from 1.6 °C to 1.0 °C, a value more comparable to the instrument accuracy of the instrument (0.5 °C). Strikingly, the removal of three suspect data sets from the “after” data reduced their standard deviation from 0.6 °C to 0.5 °C, equal to the instrument accuracy. This is a good indication that practically all systematic error has been eliminated from the quality controlled, “after” data set. The mean of the “before” data decreased by 0.4 °C, bringing it more in line with the “after” data (Tables 1 and 2). The elimination of systematic error decreased the standard deviation of the data by 1.0 °C, while quality control only decreased it by 0.5 °C. Therefore, we accept Hypothesis 3.

For completeness, we also recalculated the Student’s T-test statistics, yielding a T-statistic of 14.7 and a p-value on the order of 10^{-30}, indicating that our certainty in rejecting the null hypothesis H_0 actually *increased* when the suspect data were removed. These results reconfirm Hypotheses 1 and 2, while also providing evidence in favor of Hypothesis 3.

Some discussion of experimental limitations and assumptions is in order.

Assumption 1: Students did not have *a priori* working knowledge of how to read *meteorological* liquid-in-glass thermometers, although they may have encountered other liquid-in-glass thermometers (such as oral thermometers or chemistry lab thermometers) in the past. The effects of instruction on the measurement variance (Fig. 1 and Table 1) suggests that this assumption was valid. Additionally, distributions of temperature were plotted for all 21 participants (Fig. 4). Interquartile ranges for all 21 participants overlapped in the range (24.5 °C, 27.0 °C). A few participants (AV, JH, and SB) exhibited relatively small interquartile ranges (< 1.0 °C), indicating that they may have had prior experience reading liquid-in-glass thermometers in a scientific setting. However, the vast majority of the participants exhibited interquartile ranges >1.0 °C, validating Assumption 1 for the group as a whole.

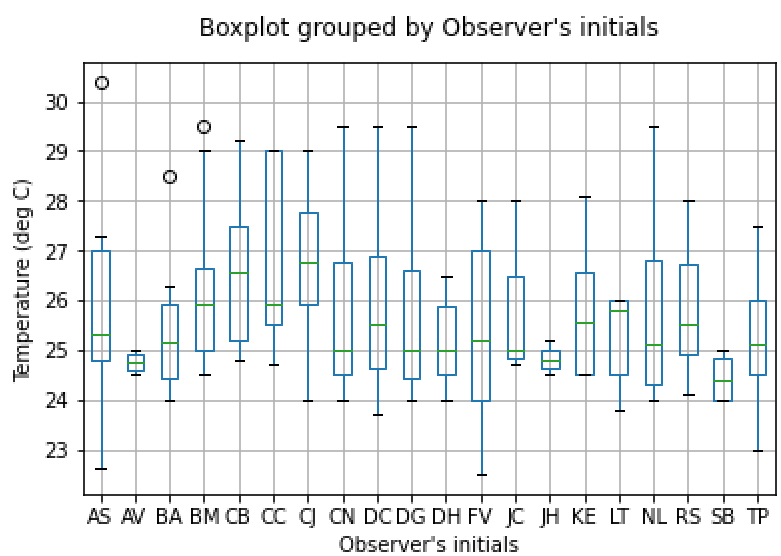
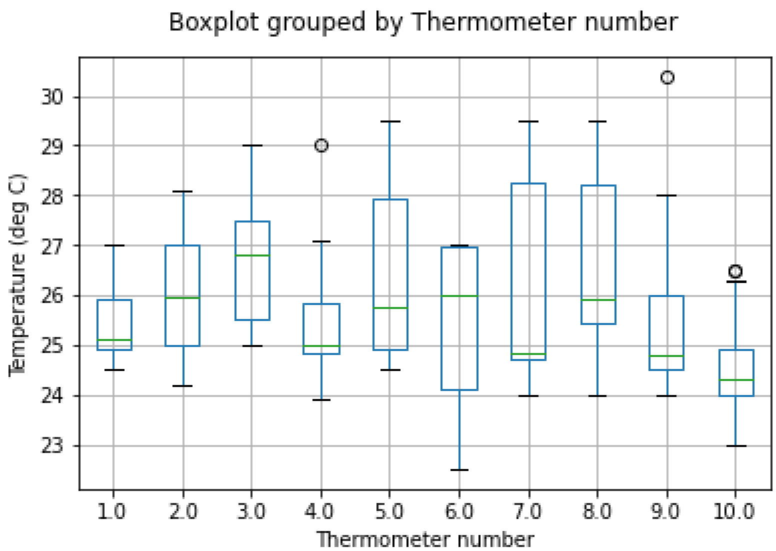


Fig. 4. As in Fig. 2, but grouped by observer (denoted by their first and last initials).

Assumption 2: The temperature in the classroom stayed constant throughout the data collection, because the classroom remained enclosed, the number of students in the room did not change during data collection, and the thermostat was not adjusted during the experiment. It is conceivable that the building HVAC system might have turned on and off during the experiment, circulating and cooling the air in the room by a few degrees Celsius, but this was not noted by the instructor (nor any of the participants). The “true” temperature in the classroom during the experiment is unknown; no additional instruments were used to measure it or any changes during the experiment.

Assumption 3: All 10 thermometers were in good working order. The thermometers were acquired in 2016 and are well within their expected useful life of 10 years. At least 12 measurements were taken with each thermometer during the experiment. The interquartile ranges for all 10 thermometers overlap in the range (24.5 °C, 27.0 °C). The only thermometers that look like they may be out of spec are Thermometers 6 and 10, which consistently exhibited lower measurements than the other nine. However, neither was deviant enough to warrant replacement or elimination from the set.



As in Fig. 2, but grouped by thermometer number (1-10).

Conclusions:

We demonstrated the importance of proper meteorological measurement technique by performing repeated measurements of air temperature in a classroom using 10 liquid-in-glass thermometers. Students collected measurements for five minutes without specific instructions, then repeated the measurements after receiving specific instructions for thermometer reading. The result was a contraction in the variance of the temperature measurements and convergence on a likely correct value of temperature.

We conclude the following points with high confidence:

- **The temperature in SC 231 at 10:15 a.m. on 25 August 2021 was 24.7 °C ± 0.5 °C.**
- The instruction in proper measurement technique nearly eliminated all systematic error from the measurements, as evidenced by the reduction in standard deviation to a value near the reported accuracy of the thermometers (0.5 °C). (This value represents the impact of random error only.)
- A smaller benefit was realized from quality control than from elimination of systematic error. However, the combination of both reduced the standard deviation of the temperature measurements to equal the reported accuracy of the thermometers.

We conclude that systematic error should be eliminated from atmospheric measurements whenever possible. While quality control can reduce the effects of systematic error, eliminating systematic error before measurements are collected is far more beneficial and effective.

Acknowledgments:

The thermometers were provided by the EAPS department. Software used in the preparation of this report includes Python 3.6, Matplotlib 3.4.3, and Pandas 1.3.2, and Scipy 1.7.1.

Appendix A: Traceability statement for the Easy-Read Thermometers.

This scanned leaflet was enclosed with each thermometer by the manufacturer, indicating instrument accuracy and affirming NIST-traceability.

Traceability Statement

Easy-Read® Thermometers

Accuracy for a thermometer with a range between -20 (-14°F) and 150°C (302°F) is ± :5 scale division; ±1 scale division below 0°C and above 70°C (158°F) and ± 1.5 scale divisions above 110°C (230°F). Accuracy for a thermometer with a range between -35 (-31°F) and 70°C (158°F) is ± 1 scale division. Accuracy for a thermometer with a range between -10 (14°F) and 260°C (500°F) is ± 1 scale division; ± 1.5 scale divisions above 100°C (230°F); ± 2 scale divisions above 200°C (392°F) and ± 2.5 scale divisions above 230°C (446°F). This thermometer was manufactured and tested against thermometer standards traceable to the National Institute of Standards and Technology (NIST). Our test procedures are based on NIST Special Publication 1088 and conform to ISO 17025. Each thermometer can also be individually certified at specific temperature points of your choice for an additional charge. Contact us for details.

The Standard Serial Number is based on the range of the thermometer. The Standard Serial Numbers calibrated by NIST are as follows: #7713700 (NIST) for ranges below -30°C, #844016 (NIST) for ranges from -30°C to 10°C, #878708 (NIST) for ranges from 0°C to 50°C, #9810984(NIST) for ranges from 50°C to 100°C, #905354 (NIST) for ranges from 100°C to 150°C, #878735 (NIST) for ranges from 150°C to 200°C #8611151 (NIST) for ranges from 200°C to 250°C, #878758 (NIST) for ranges from 250°C to 300°C

The H-B Instrument laboratory is accredited in accordance with the recognized International Standard ISO/IEC 17025:2005 through A2LA. H-B's laboratory also meets the requirements of ANSI/NCSL Z540-1-1994. The expanded measurement uncertainties associated with our calibration system are: ±0.074°C from -80°C to -1°C, ±0.041°C at the ice point in melting ice bath, ±0.045°C from 1°C to 100°C, ±0.051 from 101°C to 200°C, ±0.047°C from 201°C to 300°C, and ±0.052°C from 301°C to 400°C. These uncertainties have been calculated using our Work Instruction WI-19 to 22 that utilizes methods found in NIST Technical Note 1297. The reported uncertainty represents and expanded uncertainty expressed at approximately the 95% confidence level using a coverage factor of K=2.

Richard Jackson, Production Manager
H-B Instrument • Triple Accredited
ISO 9001:2008 Accredited Manufacturer ISO 14001:2004 Accredited Manufacturer ISO/IEC 17025:2005 Accredited Laboratory
For more information call (610) 489-5500 • Fax (610) 489-9100 • info@hbinstrument.com • www.hbinstrument.com

Form 0-656 996020002 Rev. 11

References Cited:

Kalpić, D., N. Hlupić, and M. Lovrić, 2011: Student's t-Tests. *International Encyclopedia of Statistical Science*, M. Lovric, Ed., Springer, 1559–1563.

UCAR/COMET, 2017: Foundations of Meteorological Instrumentation and Measurements. https://www.meted.ucar.edu/instrumentation/instrumentation_intro/index.htm (Accessed September 4, 2021).

Virtanen, P., and Coauthors, 2020: SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*, **17**, 261–272, https://doi.org/10.1038/s41592-019-0686-2.