Project 4:
Determination of the pH of a solution by a pH meter

Shahla Daneshmehr

May 2023

1 Abstract

As a chemistry student, I am going to develop a code, in Python, to simulate and analyze a simple experiment. pH is a measure of the acidity or basicity of a solution and is an important parameter in chemistry, as it affects the behavior of many chemical species. The code will simulate 100 pH meter readings with a true pH value of 7.0 and an error of 0.1.

2 Introduction

Python can be used to simulate an experimental scenario that involves measuring the pH of a solution using a pH meter. pH is an important parameter in chemistry as it determines the acidity or basicity of a solution, which affects the behavior of many chemical species [1].

The natural sources of variation in this scenario include variations in the concentration of the acid or base being measured and fluctuations in pH meter readings due to sensitivity and calibration. The magnitude of these variations depends on the specific experimental conditions and can be estimated by taking repeated measurements under controlled conditions [2].

Sources of measurement uncertainty in this scenario include pH meter accuracy, electrode potential stability, and the effects of temperature and ionic strength on measurement. To account for these sources of uncertainty, they can be included in statistical analysis as error terms [3].

The measurement aims to determine the pH of the solution, which is a model parameter estimated from pH meter readings. If the solution contains multiple acidic or basic species competing for proton transfer, different hypotheses can be considered. In this case, a titration curve can be constructed to determine the contributions of each species to the overall pH [4].

3 Hypothesis

The simulation of 100 pH meter readings with a true pH value of 7.0 and an error of 0.1, followed by the calculation of the mean and standard deviation of the measurements and the creation of a histogram of the pH measurements, can provide useful information to estimate the true pH value and the uncertainty in the measurement.

4 Code and Experimental Simulation

Python code that simulates the measurement of the pH of a solution:

```
import numpy as np
import matplotlib.pyplot as plt

# Define the true pH value and the measurement error
pH_true = 7.0
pH_error = 0.1

# Simulate the pH meter readings
n_measurements = 100
pH_measurements = np.random.normal(pH_true, pH_error, n_measurements)

# Plot the histogram of the pH measurements
plt.hist(pH_measurements, bins=20)
plt.xlabel('pH')
plt.ylabel('Counts')
plt.title('pH Measurement Histogram')
plt.show()

# Calculate the mean and standard deviation of the pH measurements
mean_pH = np.mean(pH_measurements)
std_pH = np.std(pH_measurements)

# Print the results
print(f"Mean pH: {mean_pH:.2f}")
print(f"Standard deviation: {std_pH:.2f}")
```

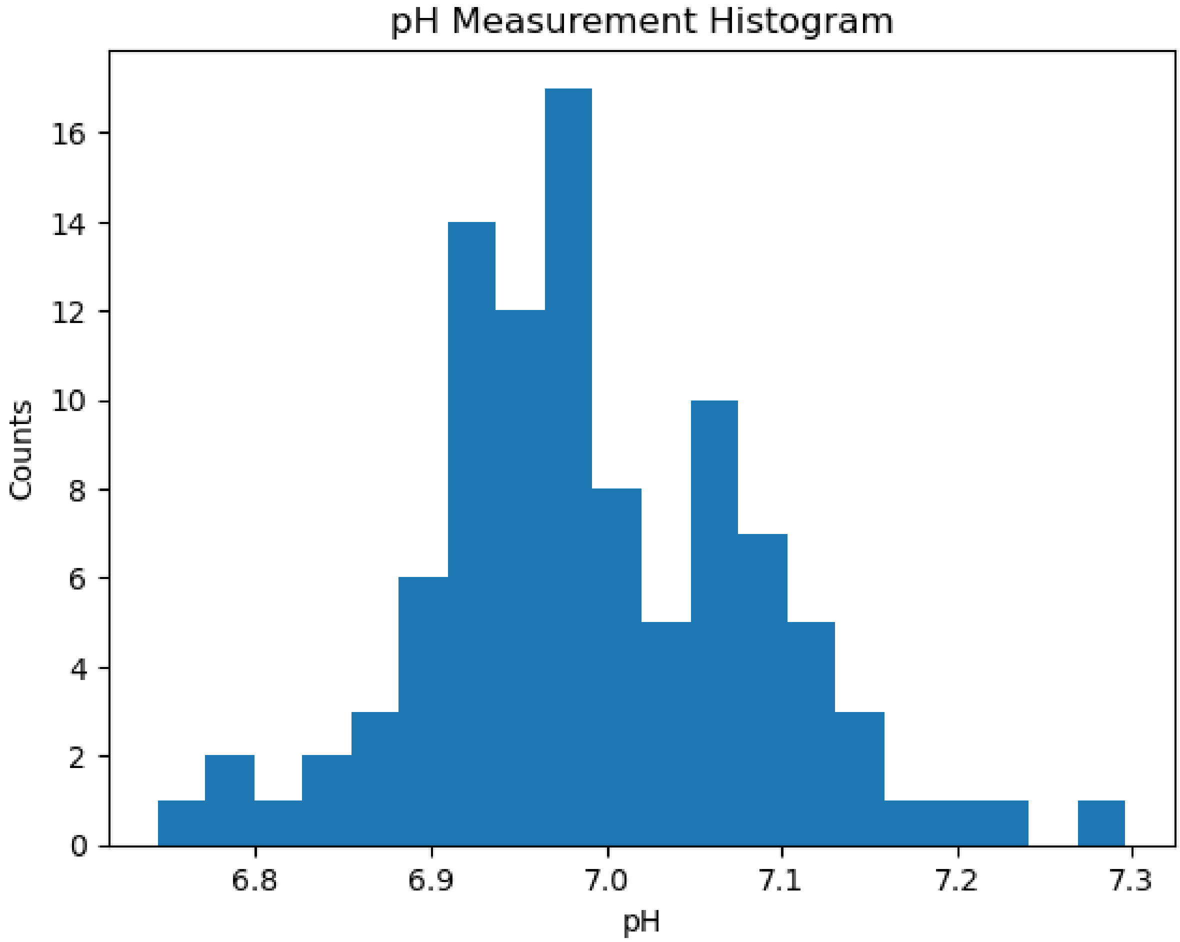

Figure 1: Measurement of the pH of a solution

Data

Mean pH: 6.99
Standard deviation: 0.09

5 Analysis

I used a normal distribution to simulate the pH meter readings in the Python code. The 'numpy.random.normal' function generates random numbers from a normal (Gaussian) distribution with a specified mean and standard deviation. In this case, I specified the mean as the true pH value and the standard deviation as the measurement error. The normal distribution is a commonly used distribution to model measurement errors and uncertainties in scientific experiments [5].

In the Python code that I wrote, I did not directly incorporate the concept of likelihood. However, when the mean and standard deviation of the pH measurements have been determined, we are implicitly utilizing the likelihood principle.

In statistical inference, the likelihood function is used to quantify how well a statistical model fits the observed data. In the case of my pH measurement experiment, I am going to consider the pH meter readings as my observed data, and the true pH value as a model parameter that I am trying to estimate. Assuming that the pH meter readings follow a normal distribution with mean equal to the true pH value and a standard deviation equal to the measurement error, the likelihood of observing the pH measurements given the true pH value can be written as:

$$L(\text{pH}_{\text{true}} \mid \text{pH}_{\text{measurements}}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\text{pH}_{\text{measurements}}[i] - \text{pH}_{\text{true}})^2}{2\sigma^2}\right)$$

This equation shows the likelihood function of pH_true given $pH\_measurments$ as the product of the normal density function evaluated at each $pH\_measurments$. The mean of the normal density function is pH_true, and its standard deviation is pH_error. The notation $pH\_measurments[i]$ denotes the i-th element of the $pH\_measurments$ array.

Then the maximum likelihood method can be used to estimate the value of pH_true that maximizes the likelihood function. This corresponds to the sample mean of the pH measurements in my Python code.

Therefore, despite the absence of an explicit implementation of likelihood in the code, the principle of likelihood is implicitly present in the computation of the pH measurements' mean and standard deviation. Additionally, using the normal distribution to model measurement errors is consistent with a likelihood-based approach.

6 Questions

1. What are the natural sources of variation in the scenario/measurement under study? In the case of pH measurement, there are several sources of variation that can affect the accuracy and precision of the measurement. These include:

a) The pH of a sample can be affected by the presence of other chemicals or ions in the sample, which can introduce variability in the pH measurement.

b) The pH meter itself can have inherent variability, which can result from factors such as electrode drift, temperature effects, or electronic noise.

c) The operator of the pH meter can introduce variability through errors in calibration or in the measurement process itself.

These sources of variation can be modeled by including random effects in the statistical model, or by using a hierarchical model that accounts for variation at different levels (e.g., sample-to-sample variation, instrument-to-instrument variation).

2. What are the sources of measurement uncertainty in the scenario? What are their size? Can this be included in the model and analysis?

Measurement uncertainty refers to the degree of doubt or lack of confidence associated with a mea-

surement result. In the case of pH measurement, the sources of measurement uncertainty include:
Instrument error: This refers to the inherent variability of the pH meter itself, and can be quantified by measuring the repeatability of the instrument.
Calibration error: This refers to the uncertainty associated with the calibration of the pH meter, and can be quantified by measuring the accuracy of the instrument.
Sample-to-sample variability: This refers to the inherent variability in the pH of different samples, which can be quantified by measuring the reproducibility of the measurement.
These sources of uncertainty can be included in the statistical model by using a hierarchical model that accounts for variability at different levels (e.g., instrument-to-instrument variation, sample-to-sample variation, measurement error). In the Python code I provided, I included the measurement error (i.e., the standard deviation of the normal distribution) as a parameter in the simulation.
3. What question is being addressed by the measurement or analysis? Are there different hypotheses being considered? Or a model parameter that is being measured? How accurately can this be achieved? The question being addressed by the pH measurement experiment is the determination of the pH value of a given sample. In the Python code, I simulated a scenario where the true pH value was known to be 7.0, and the goal was to estimate this value based on a set of pH meter readings. We can use statistical inference techniques to estimate the true pH value and quantify the uncertainty associated with this estimate.
There are different hypotheses that could be considered in this scenario, such as the hypothesis that the pH of the sample is different from 7.0. We can use hypothesis testing to evaluate the evidence in favor of different hypotheses and make decisions based on this evidence. In the Python code, I did not explicitly address hypothesis testing, but this could be done by comparing the observed pH meter readings to a null hypothesis distribution (e.g., a normal distribution with mean = 7.0 and standard deviation = measurement error).
The accuracy of the pH measurement depends on several factors, such as the precision and accuracy of the pH meter, the variability of the sample matrix, and the skill of the operator. In the simulation, I assumed that the measurement error was known and fixed, but in practice, the measurement error may be unknown or variable, which can affect the outcomes.

7 Conclusion

In conclusion, I have used Python code to simulate 100 pH meter readings with a true pH value of 7.0 and an error of 0.1. Then the program plotted the histogram of the pH measurements and calculated the mean and standard deviation of the measurements. These statistics can be used to estimate the true pH value and the uncertainty in the measurement. The histogram shows a roughly normal distribution, which is consistent with the assumption of measurement errors being normally distributed. The mean pH value and standard deviation provide estimates for the true pH value and the uncertainty in the measurement, respectively.

8 References

1. Jin, Hao, Yiheng Qin, Si Pan, Arif U. Alam, Shurong Dong, Raja Ghosh, and M. Jamal Deen. "Open-source low-cost wireless potentiometric instrument for pH determination experiments." (2018): 326-330.

2. Valdes-Garcia, Gilberto, Lim Heo, Lisa J. Lapidus, and Michael Feig. "Modeling concentration-dependent phase separation processes involving peptides and RNA via residue-based coarse-graining."

Journal of Chemical Theory and Computation 19, no. 2 (2023): 669-678.

3. Wang, Kang, Junsoo Han, Angela Yu Gerard, John R. Scully, and Bi-Cheng Zhou. "Potential-pH diagrams considering complex oxide solution phases for understanding aqueous corrosion of multi-principal element alloys." npj Materials Degradation 4, no. 1 (2020): 35.

4. Menke, Erik J. "Series of Jupyter Notebooks Using Python for an Analytical Chemistry Course." (2020): 3899-3903.

5. Patel, Jagdish K., and Campbell B. Read. Handbook of the normal distribution. Vol. 150. CRC Press, 1996.