

Problem 1 (40 points)

The following data gives the average pH in rain/sleet/snow for the two-year period 2004–2005 at 20 rural sites on the U.S. West Coast. (Source: National Atmospheric Deposition Program):

5.335	5.345	5.395	5.305	5.315
5.380	5.520	5.190	5.455	5.330
5.360	6.285	5.350	5.125	5.115
5.510	5.340	5.340	5.305	5.265

For all calculations, you can use the sample mean and the sample variance calculated from R. You can create a data in R with the following R script:

```
ph <- c(5.335, 5.345, 5.395, 5.305, 5.315, 5.380, 5.520, 5.190, 5.455, 5.330,
        5.360, 6.285, 5.350, 5.125, 5.115, 5.510, 5.340, 5.340, 5.305, 5.265)
```

Note: Please use 4 decimal digits in your solution files and in your calculation.

- (1) **(3 points)** Find the sample mean, sample variance, and sample standard deviation.
- (2) **(11 points)** Perform a t -test to see if the mean pH is at least 5.40 with the significance level of 0.05. Please clearly specify 5 steps used in the test and use the rejection approach. You need to use the t -table to find t_α or $t_{\frac{\alpha}{2}}$.
- (3) **(5 points)** Perform a t -test based on the confidence interval approach. You only need to calculate the corresponding confidence interval and present Step 5 for this part. You need to use the t -table to find t_α or $t_{\frac{\alpha}{2}}$.
- (4) **(2 points)** Use R to calculate the p -value for this test from (2).
- (5) **(4 points)** Find the two-sided 95% confidence interval for the pH. Use the t -table to find t_α or $t_{\frac{\alpha}{2}}$.
- (6) **(5 points)** Draw the QQ-plot and then comment on its normality with one or two sentences.
- (7) **(5 points)** Draw box plot of the pH and identify any extreme values. Would the sample mean or the sample median be a better descriptor of typical pH values?
- (8) **(5 points)** Remove the most extreme value, then perform a t -test to see if the mean pH is at least 5.40 with the significance level of 0.05. You only need to present your p -value and Step 5 for this part.