Problem 1

37/40

1.1

Sample mean = 5.3782

Sample Variance = 0.0564

Sample Standard Deviation = 0.2375

1.2

Step 1: Hypothesis

Null Hypothesis (H_o) : μ = 5.40

-2

Alternate Hypothesis (H_a) : μ < 5.40

Step 2: Test Statistic

α = 0.05, n = 20,

t = (xbar - μ) / (s / sqrt(n))

with degrees of freedom df = n - 1 = 19

t = (xbar - 5.40) / (0.0564 / sqrt(20))

Step 3: Rejection region

For a one-tailed (lower-tail) test at α = 0.05 with 19 df, the critical value from the t-table is -1.7291

Step 4: Compute test statistic

t = (5.3782 - 5.40) / (0.0564 / sqrt(20))

t = -0.4096

Step 5: Decision and Interpretation

Since -0.4096 is not less than -1.7291

We fail to reject the null hypothesis H_o

At the 0.05 significance level, data provides insufficient evidence to conclude that the mean pH is less than 5.40

1.3

For the confidence interval approach we should construct a one-sided 95% confidence interval for $\boldsymbol{\mu}$ and then check whether 5.40 is contained in this interval.

The 95% confidence interval (CI) is given by:

$$(-\infty, x + t_{0.025,\, df} * s / \sqrt{n})$$

Lower bound = $-\infty$
Upper bound is $x + t_{0.025,\, df} * s / \sqrt{n}$

$t_{0.025,\, 19}$ is the critical value for two-tail test with $\boldsymbol{\alpha} = 0.05$, df = 19 , by using R we get

$$5.37825 + 1.729133 * (0.2374884 / 20)$$

CI lower = $- \infty$

CI upper = 5.4701

Step 5:

95% Confidence Interval for μ: ($-\infty$, 5.4701). Since 5.40 is inside this interval, we fail to reject H_0.

Data provides insufficient evidence to conclude that the mean pH is less than 5.40.

1.4

p value using R

p_value <- pt(t_stat, df = df)

p_value = 0.3433

1.5

we should construct a two-sided 95% confidence interval for $\boldsymbol{\mu}$ and then check whether 5.40 is contained in this interval.

The 95% confidence interval (CI) is given by:

$$x \pm t_{0.025,\, df} * s / \sqrt{n}$$

Lower bound = $x - t_{0.025,\, df} * s / \sqrt{n}$

Upper bound = $x + t_{0.025,\, df} * s / \sqrt{n}$

$t_{0.025,\,19}$ is the critical value for two-tail test with α / 2 = 0.025, df = 19 , by using R we get

$$5.37825 \pm 5.37825 * (0.2374884 / 20)$$

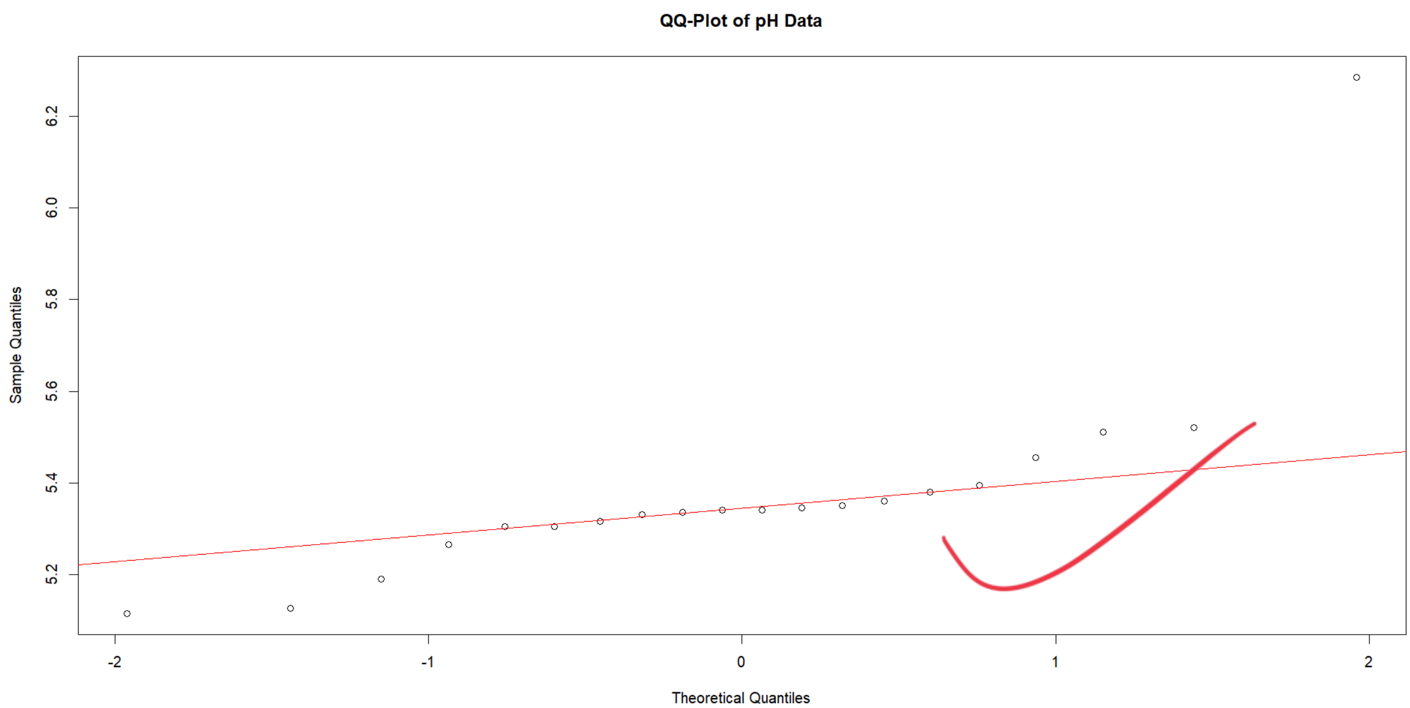
CI lower = 5.2671

CI upper = 5.4701

95% Confidence Interval for μ: (5.2671 , 5.4701)

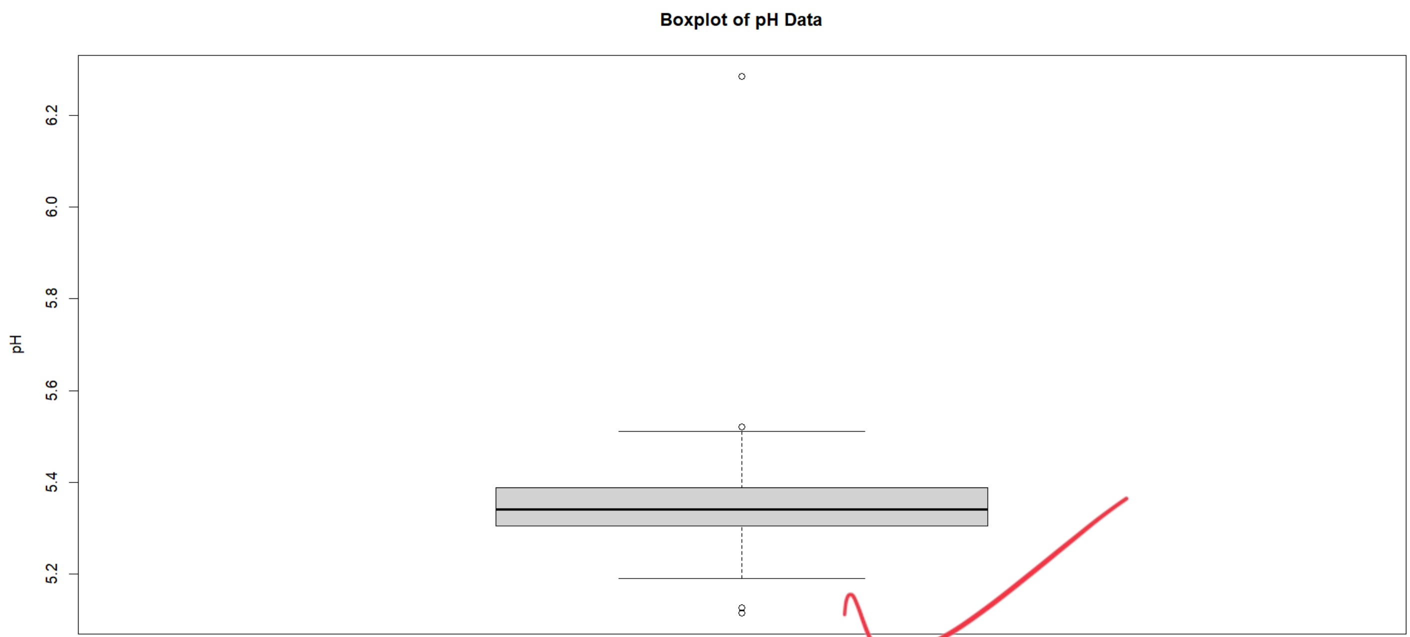
1.6

QQ plot



The QQ-plot appears approximately linear, which suggests that the pH data is roughly normally distributed. Although, there appears to be one outlier which can influence the normality.

1.7



From the box plot we can see that there appears to be three outliers. The median is below 5.4. Most values lie in the range between 5.3 to 5.4 indicating low variance.

4 of them, specify them
and 6.285 is an extreme outlier

-1

1.8

outlier = 6.285, after removing

Sample mean = 5.3305

Sample variance = 0.0115

Sample standard deviation = 0.107

n = 19

t = (xbar - μ) / (s / sqrt(n))

t = (5.3305 - 5.40) / (0.107   / sqrt(19))

t = -2.8297

p_value = 0.0056

Because the p-value = 0.0056 is well below 0.05, we reject the null hypothesis at the 5% significance level. After removing the outlier, there is strong evidence that the true mean pH is less than 5.40.

```
# R code
# pH data
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)

# Sample size
n <- length(ph)

# (1)
xbar <- mean(ph)
s2   <- var(ph)
s    <- sd(ph)

#  rounded to 4 decimal places
cat("Sample Mean =", round(xbar, 4), "\n")
cat("Sample Variance =", round(s2, 4), "\n")
cat("Sample Standard Deviation =", round(s, 4), "\n")

t_stat <- (xbar - 5.40) / (s / sqrt(n))
cat("t-statistic =", round(t_stat, 4), "\n")


df <- n - 1
t_crit <- qt(0.05, df = df)
cat("t-critical =", round(t_crit, 4), "\n")


alpha <- 0.05
t_val <- qt(0.95, df = df)
margin <- t_val * (s / sqrt(n))
CI_lower <- -Inf
CI_upper <- xbar + margin
cat("95% Confidence Interval for μ: (", round(CI_lower, 4), ",", round(CI_upper, 4), ")\n")

# p_value <- pt(t_stat, df = df)
cat("One-tailed p-value =", round(p_value, 4), "\n")


# two sided
alpha <- 0.05
t_val <- qt(1 - alpha/2, df = df)
margin <- t_val * (s / sqrt(n))
```

```r
CI_lower <- xbar - margin
CI_upper <- xbar + margin
cat("95% Confidence Interval for μ: (", round(CI_lower, 4), ",", round(CI_upper, 4), ")\n")


t_val <- qt(1 - alpha/2, df = df)
margin <- t_val * (s / sqrt(n))
CI_lower <- xbar - margin
CI_upper <- xbar + margin
cat("95% Confidence Interval for μ: (", round(CI_lower, 4), ",", round(CI_upper, 4), ")\n")


qqnorm(ph, main = "QQ-Plot of pH Data")
qqline(ph, col = "red")

boxplot(ph, main = "Boxplot of pH Data", ylab = "pH")


ph_no_outlier <- ph[ph != 6.285]
n_no <- length(ph_no_outlier)
xbar_no <- mean(ph_no_outlier)
s_no <- sd(ph_no_outlier)
df_no <- n_no - 1

t_stat_no <- (xbar_no - mu0) / (s_no / sqrt(n_no))
p_value_no <- pt(t_stat_no, df = df_no)

cat("After removing the outlier:\n")
cat("Sample Mean =", round(xbar_no, 4), "\n")
cat("t-statistic =", round(t_stat_no, 4), "\n")
cat("p-value =", round(p_value_no, 4), "\n")
```