Bivariate statistics t-test

Andri Setiyawan Benedikt Meyer Niri Gala
Yosep Dwi Kristanto

Problem

Use the data file `SPSS_Gapminder_main_1.sav`
Run t-tests correctly and check assumptions carefully.

1. Test the life expectancy for the population in all countries and compare male and female scores.
2. Compare the life expectancy of males and females in OECD, G77, and Other countries separately.

Use the data file `Pollution.sav`
Assume that you are an environmental lobbyist interested in assessing whether the introduction of a new waste removal system at a large chemical factory has had an impact on groundwater contamination levels at 12 test sites in the area. You compare contamination levels from the same 12 locations based on samples collected one month before, and one month after the system was introduced.

1. Run the appropriate t-test for `pollute_before` and `pollute_after` as your variables of interest
2. Was there a significant difference between the average groundwater contamination levels measured before and after the waste removal system was installed? Run appropriate t-test

Table 1: Summary statistics for male and female life expectancy across countries in 2000 and 2010

Year	Gender	M	SD
2000	Female	68.78098	10.699871
2000	Male	63.92717	9.583993
2010	Female	72.23913	9.122901
2010	Male	67.42446	8.256033

1 Male and Female Life Expectancy Across Countries

1.1 Data Exploration

Table 1 presents summary statistics for male and female life expectancy across countries in 2000 and 2010, while Figure 1 illustrates the distribution of male and female life expectancy across countries for these years.

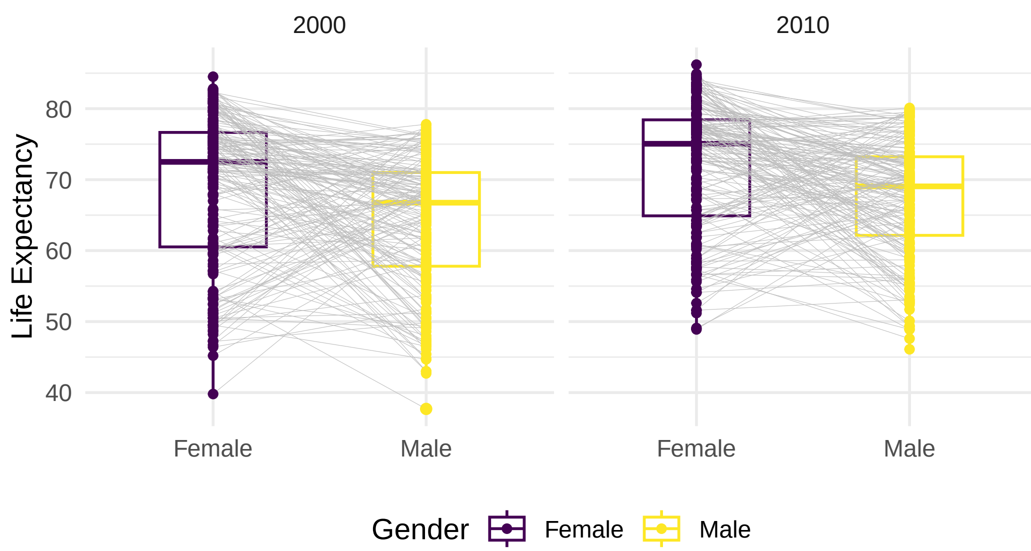


Figure 1: Comparison of male and female life expectancy across countries for the years 2000 and 2010

1.2 Hypotheses

1.2.1 2000

H_0: There is no difference in life expectancy between males and females across countries in 2000.

H_1: There is a difference in life expectancy between males and females across countries in 2000.

1.2.2 2010

H_0: There is no difference in life expectancy between males and females across countries in 2010.

H_1: There is a difference in life expectancy between males and females across countries in 2010.

1.3 Assumption Checking

- Since the variable life expectancy is a ratio variable, the first assumption is met.
- The grouping variable is dichotomous, categorized as male and female.
- The distribution of the differences in life expectancy between males and females across countries in 2000 and 2010 is presented in Figure 2.

  Figure 3 presents QQ plots illustrating the distribution of life expectancy differences between males and females across countries for the years 2000 and 2010.

  Based on Figure 3, the differences in life expectancy between males and females in both 2000 and 2010 do not align well with a normal distribution. This suggests that the assumption of normality for the life expectancy differences across these years may not hold.

```
    Shapiro-Wilk normality test

data:  gapminder_data_fm$life_exp_diff[gapminder_data_fm$year == 2000]
W = 0.9642, p-value = 0.00012
```

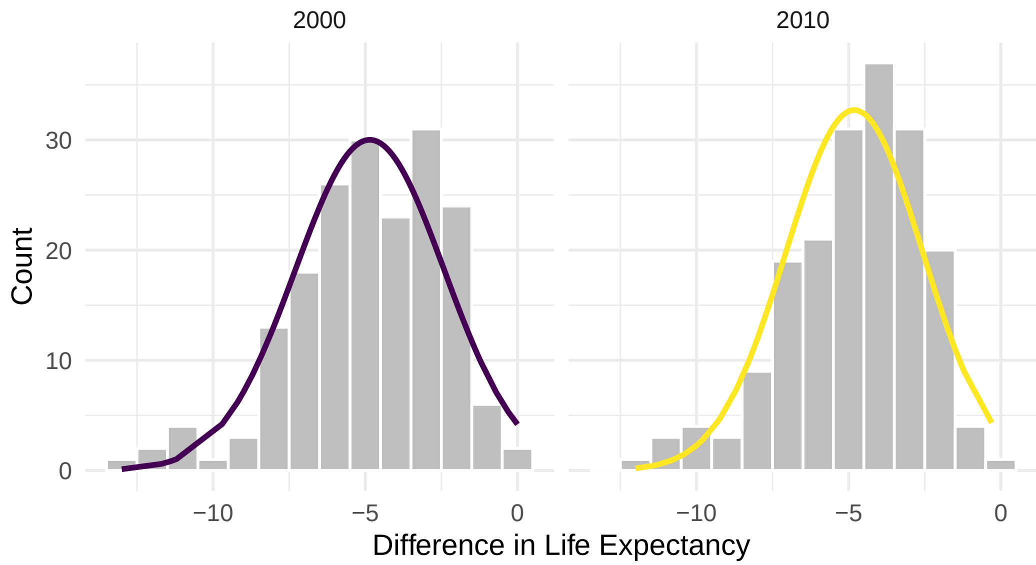


Figure 2: Distribution of differences in life expectancy between males and females across countries in 2000 and 2010

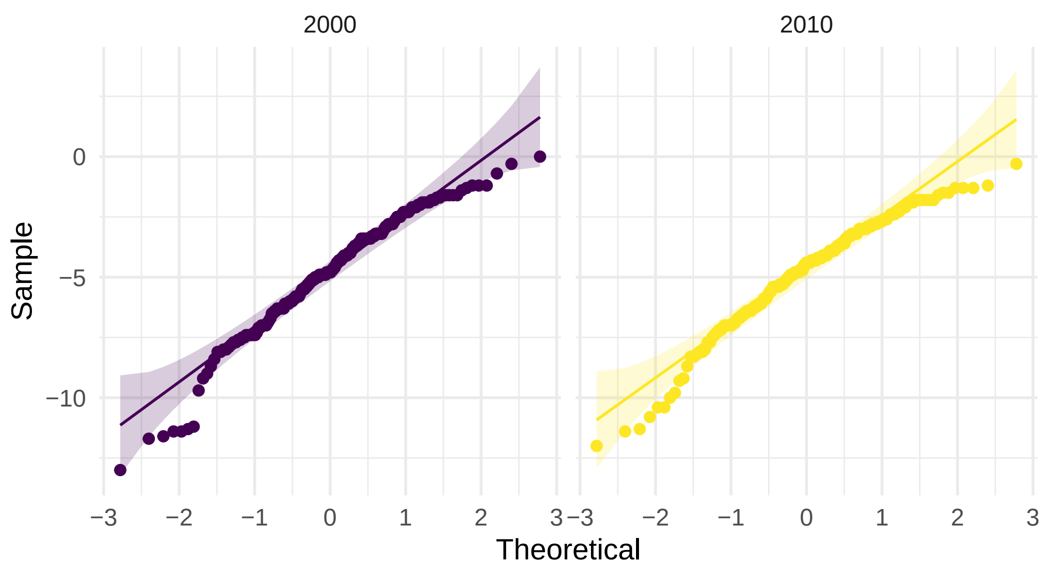


Figure 3: QQ plots displaying the distribution of differences in life expectancy between males and females across countries for the years 2000 and 2010

```
    Shapiro-Wilk normality test

data:  gapminder_data_fm$life_exp_diff[gapminder_data_fm$year == 2010]
W = 0.96217, p-value = 7.259e-05
```

To confirm the graphical findings, we conducted a Shapiro–Wilk test of normality for life expectancy differences in 2000 and 2010. The test yielded p-values of 1.2004593×10^{-4} for 2000 and 7.2592831×10^{-5} for 2010, both below .05, indicating the differences are unlikely to be from a normally distributed population.

Although the life expectancy differences for both years cannot be assumed to come from a normally distributed population, the sample size (184 for each year) is large. Based on the Central Limit Theorem, we can still proceed with a t-test.

1.4 Calculate the t Statistics

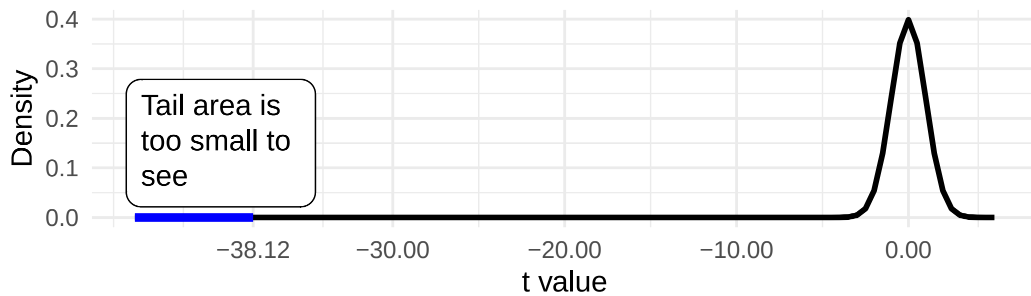
Paired t-tests were conducted to compare male and female life expectancy across countries. The t-statistics, degrees of freedom (df), and p-values for each test are presented below.

```
    Paired t-test

data:  gapminder_data_fm$LEm00 and gapminder_data_fm$LEf00
t = -38.123, df = 367, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -5.104173 -4.603436
sample estimates:
mean difference
      -4.853804


    Paired t-test

data:  gapminder_data_fm$LEm10 and gapminder_data_fm$LEf10
t = -41.22, df = 367, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -5.044365 -4.584983
sample estimates:
mean difference
      -4.814674
```

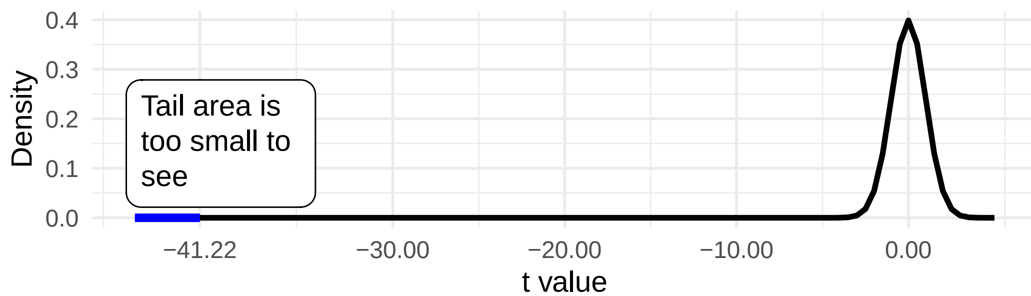
1.5 Testing for the Significance of t

Based on the results of the paired t-test for the 2000 data, we obtained a p-value of less than 2.2e-16 (see Figure 4a). This indicates that, assuming the null hypothesis is true, the probability of obtaining a sample like that in `SPSS_Gapminder_main_1 (1).sav` is less than $2.2 \cdot 10^{-16}$.

Similarly, for the paired t-test for the 2010 data, the p-value was also less than 2.2e-16 (see Figure 4b). This suggests that, assuming the null hypothesis is true, the probability of obtaining a sample like that in `SPSS_Gapminder_main_1 (1).sav` is less than $2.2 \cdot 10^{-16}$.



(a) Tail area representing p-value from paired t-test for 2000 data



(b) Tail area representing p-value from paired t-test for 2010 data

Figure 4: Theoretical t-distribution with 367 degrees of freedom

1.6 Interpreting t

From the paired t-test for the 2000 data, we obtained t = -38.1228284 and a p-value less than 2.2e-16. Since the p-value is smaller than .05, we reject the null hypothesis.

Similarly, for the 2010 data, the paired t-test yielded t = -41.2197962 and a p-value less than 2.2e-16. As the p-value is also smaller than .05, we reject the null hypothesis.

1.7 Effect Size

We computed the effect size of the paired *t*-test using Pearson's r and Cohen's d. Table 2 presents the values of r and d for both the 2000 and 2010 data.

```
# Calculating r and d for 2000 data
r_life_exp_all00 <- sqrt(
  as.numeric(t_test00$statistic)^2 / (as.numeric(t_test00$statistic)^2 + as.numeric(t_test00$
)
d_life_exp_all00 <- 2 * as.numeric(t_test00$statistic) / sqrt(as.numeric(t_test00$parameter))

# Calculating r and d for 2010 data
r_life_exp_all10 <- sqrt(
  as.numeric(t_test10$statistic)^2 / (as.numeric(t_test10$statistic)^2 + as.numeric(t_test10$
)
d_life_exp_all10 <- 2 * as.numeric(t_test10$statistic) / sqrt(as.numeric(t_test10$parameter))
```

Table 2: Effect size for 2000 and 2010 data

Effect Size	2000	2010
Pearson's r	0.893527	0.9068452
Cohen's d	-3.9799909	-4.3033117

1.8 Report the Findings

The results of the paired *t*-test for the 2000 data revealed a significant difference in life expectancy between males (M = 63.93, SD = 9.58) and females (M = 68.78, SD = 10.7), t(367) = -38.12, p < .001. The effect size was large, with Pearson's r = 0.89 and Cohen's d = -3.98, according to Cohen's (1988) guidelines.

Similarly, the paired *t*-test for the 2010 data showed a significant difference in life expectancy between males (M = 67.42, SD = 8.26) and females (M = 72.24, SD = 9.12), t(367) = -41.22, p < .001. The effect size was also large, with Pearson's r = 0.91 and Cohen's d = -4.3, indicating a stronger effect in 2010 compared to 2000.

Table 3: Summary statistics of male and female life expectancy in OECD, G77, and other countries for 2000 and 2010

Group	Year	Gender	n	M	SD
OECD	2000	Female	31	80.37419	2.247809
OECD	2000	Male	31	74.29032	2.911970
OECD	2010	Female	31	82.50645	2.001322
OECD	2010	Male	31	77.17742	2.714493
G77	2000	Female	127	65.03386	10.343407
G77	2000	Male	127	61.04409	9.421368
G77	2010	Female	127	69.01575	8.670516
G77	2010	Male	127	64.81654	7.793452
others	2000	Female	26	73.26154	5.823063
others	2000	Male	26	65.65385	6.027486
others	2010	Female	26	75.74231	5.409042
others	2010	Male	26	68.53462	5.495812

2 Male and Female Life Expectancy in OECD, G77, and Other Countries

2.1 Data Exploration

Table 3 provides summary statistics for male and female life expectancy in OECD, G77, and other countries for 2000 and 2010. The table indicates a substantial difference in life expectancy between males and females across each country group (OECD, G77, and others) in both 2000 and 2010. To determine whether this difference is systematic, we will conduct a *t*-test.

2.2 Assumption Checking

All assumptions for conducting a *t*-test are satisfied: (a) the outcome variables, male and female life expectancy, are ratio-level; (b) the grouping variable, gender, is dichotomous (male and female); and (c) each group has a sufficiently large sample size (31 for OECD, 127 for G77, and 26 for other countries).

For the group with the smallest sample size (i.e., "others"), we can assume it is drawn from a normally distributed population. The Shapiro-Wilk test of normality for this group for both 2000 and 2010 data is as follows.

```
    Shapiro-Wilk normality test
```

Table 4: Paired t-test results for life expectancy differences between males and females

Group	Year	df	t	p
OECD	2000	30	-27.61413	6.9e-23
OECD	2010	30	-22.99121	1.32e-20
G77	2000	126	-22.33942	1.2e-45
G77	2010	126	-24.16802	3.83e-49
others	2000	25	-13.70816	3.94e-13
others	2010	25	-13.58182	4.84e-13

```
data:  gapminder_data_fm$life_exp_diff[gapminder_data_fm$year == 2000 & gapminder_data_fm$OEC
W = 0.92445, p-value = 0.05722


    Shapiro-Wilk normality test

data:  gapminder_data_fm$life_exp_diff[gapminder_data_fm$year == 2010 & gapminder_data_fm$OEC
W = 0.959, p-value = 0.3723
```

Since the p-values for both tests are greater than .05, we fail to reject the assumption that the samples come from a normally distributed population.

2.3 Calculate the t Statistics

Table 4 presents the t-statistics from paired t-tests for OECD, G77, and other countries in 2000 and 2010.

2.4 Testing for Significance of t

All the p-values in the table are less than 0.05, indicating that, assuming the null hypothesis is true, the probability of obtaining a sample like ours is very small (less than 5%).

2.5 Interpreting t

Since the p-values are less than 0.05, we reject the null hypothesis. There is sufficient evidence to support the claim that male and female life expectancy differs across OECD, G77, and other countries in both 2000 and 2010.

Table 5: Effect sizes (Pearson's r and Cohen's d) for differences in male and female life expectancy across OECD, G77, and other countries in 2000 and 2010

Group	Year	df	t	p	Pearson's r	Cohen's d
OECD	2000	30	-27.61413	6.9e-23	0.9808909	-10.083255
OECD	2010	30	-22.99121	1.32e-20	0.9727764	-8.395202
G77	2000	126	-22.33942	1.2e-45	0.8935413	-3.980307
G77	2010	126	-24.16802	3.83e-49	0.9069501	-4.306116
others	2000	25	-13.70816	3.94e-13	0.9394581	-5.483265
others	2010	25	-13.58182	4.84e-13	0.9384289	-5.432728

Table 6: Summary statistics of groundwater contamination levels before and after the waste removal system installation

Time	n	M	SD
after	12	11.5	2.430862
before	12	16.5	2.067058

2.6 Effect Size

Table 5 shows effect sizes (Pearson's r and Cohen's d) for differences in male and female life expectancy across OECD, G77, and other countries for the years 2000 and 2010.

2.7 Report the Findings

The paired *t*-tests for male and female life expectancy across OECD, G77, and other countries in 2000 and 2010 revealed significant differences in each group. For OECD countries, life expectancy differences in 2000 ($t(30) = -27.6$, $p < .001$, $r = .98$, $d = -10.1$) and 2010 ($t(30) = -23.0$, $p < .001$, $r = .97$, $d = -8.4$) showed large effect sizes. Similarly, in G77 countries, 2000 ($t(126) = -22.3$, $p < .001$, $r = .89$, $d = -3.98$) and 2010 ($t(126) = -24.2$, $p < .001$, $r = .91$, $d = -4.31$) indicated large effects. For other countries, 2000 ($t(25) = -13.7$, $p < .001$, $r = .94$, $d = -5.48$) and 2010 ($t(25) = -13.6$, $p < .001$, $r = .94$, $d = -5.43$) also showed large effect sizes.

3 Groundwater Contamination Level

3.1 Data Exploration

Table 6 presents the summary statistics of groundwater contamination levels before and after the installation of the waste removal system.

Figure 5 displays the distribution of groundwater contamination levels before and after the installation of the waste removal system.

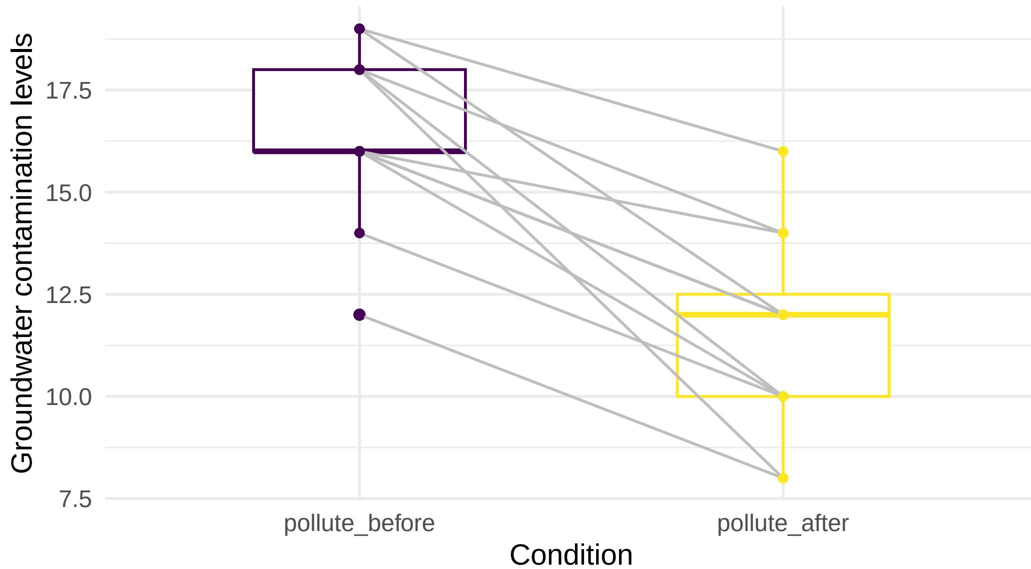


Figure 5: Distribution of groundwater contamination levels before and after the waste removal system installation

3.2 Hypotheses

H_0: There is no difference between the groundwater contamination levels measured before and after the waste removal system was installed.

H_1: There is a difference between the groundwater contamination levels measured before and after the waste removal system was installed.

3.3 Assumption Checking

- Since the variable `pollute_before` and `pollute_after` is a ratio variable, the first assumption is met.
- The grouping variable is dichotomous, categorized as before and after.
- The figure shows the distribution of differences in groundwater contamination levels before and after.

  Figure 7 shows Q-Q plot of groundwater contamination level differences before and after intervention.

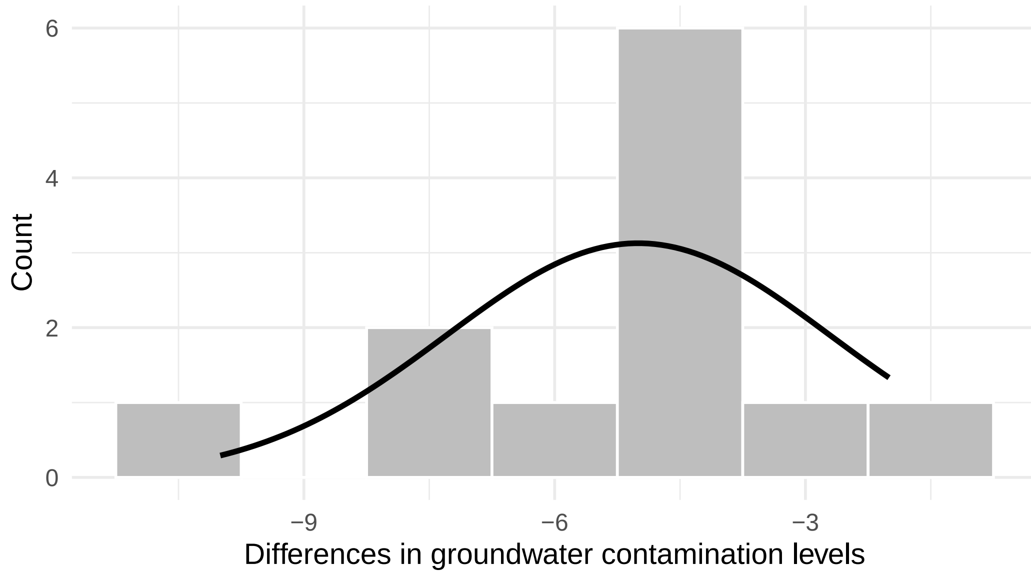


Figure 6: Distribution of groundwater contamination level differences before and after intervention

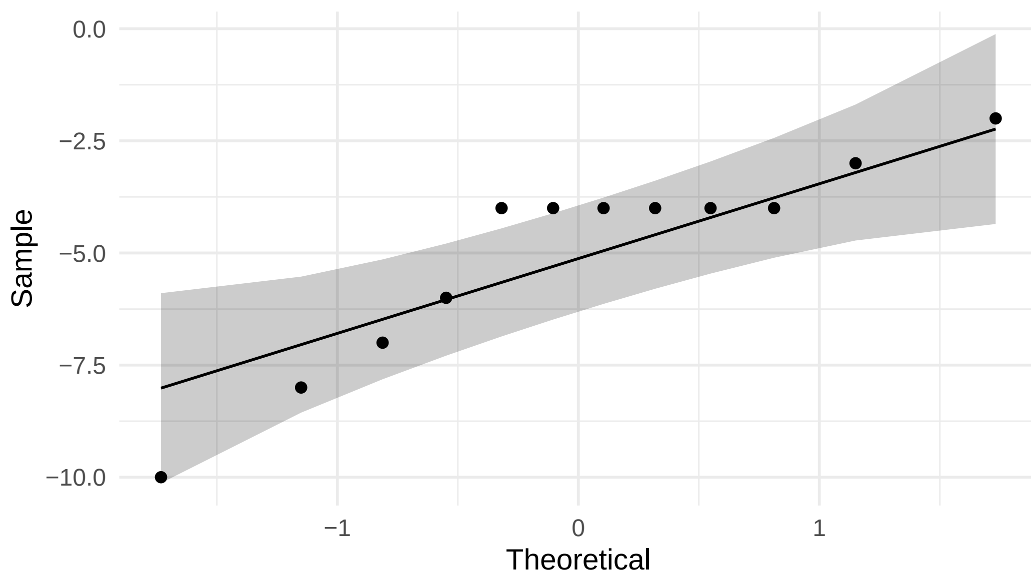


Figure 7: Q-Q plot of groundwater contamination level differences before and after intervention

Based on the Q-Q plot in Figure 7, it is unlikely that the sample comes from a normally distributed population. This observation is further supported by the following Shapiro-Wilk test, which yielded a p-value of .04927, less than the .05 threshold. Therefore, we reject the assumption that the sample is from a normally distributed population.

```
    Shapiro-Wilk normality test

data:  pollution_data$pollute_diff
W = 0.86027, p-value = 0.04927
```

Since the sample does not come from a normally distributed population, one of the assumptions for conducting a *t*-test is not met. Therefore, we will use the alternative, the Wilcoxon signed-rank test.

3.4 Calculate the Test Statistics

```
    Wilcoxon signed rank test with continuity correction

data:  pollution_data$pollute_before and pollution_data$pollute_after
V = 78, p-value = 0.002201
alternative hypothesis: true location shift is not equal to 0
```

3.5 Testing for the Significance of the Test Statistics

The p-value is 0.002201, which indicates that the probability of obtaining this sample, assuming the null hypothesis is true, is very small.

3.6 Interpreting the Test Statistics

Since the p-value of 0.002201 is smaller than 0.05, we reject the null hypothesis.

3.7 Effect Size

```
t_test_pollute <- t.test(
  x = pollution_data$pollute_before,
  y = pollution_data$pollute_after,
  paired = TRUE
```

```r
)
t_pollute <- t_test_pollute$statistic |>
  as.numeric()
df_pollute <- t_test_pollute$parameter |>
  as.numeric()
r_pollute <- sqrt(
  t_pollute^2 / (t_pollute^2 + df_pollute)
)
d_pollute <- 2 * t_pollute / sqrt(df_pollute)

print(r_pollute)
```

```
[1] 0.9154173
```

```r
print(d_pollute)
```

```
[1] 4.548588
```

We obtained Pearson's r = .92 and Cohen's d = 4.55.

3.8 Reporting the Findings

A Wilcoxon signed-rank test was conducted to determine if there was a difference in groundwater contamination levels before and after the installation of the waste removal system. The results indicated a significant difference, V = 78, p = .002201, suggesting that the contamination levels were different after the system was installed. The effect size was large, with Pearson's r = .92 and Cohen's d = 4.55, indicating a large difference between the before and after measurements.