Assignment 3

EMILY BERLINGHOFF

2024-04-06

Importing the data

```
a3 <-  read.csv("assign.csv")
```

Question 1

Null hypothesis in sentence form:

The mean daily screen time for non-North American children post-pandemic is equal to or less than the pre-pandemic level.

Null hypothesis in symbolic form:

$H_0 : \mu_{post} \leq \mu_{pre}$

Alternative hypothesis in symbolic form:

$H_a : \mu_{post} > \mu_{pre}$

Question 2

To conduct a t-test to evaluate whether the mean daily screen time for non-North American children is higher than the pre-pandemic level, the data must satisfy certain model conditions for the t-test to be valid. These conditions typically include:

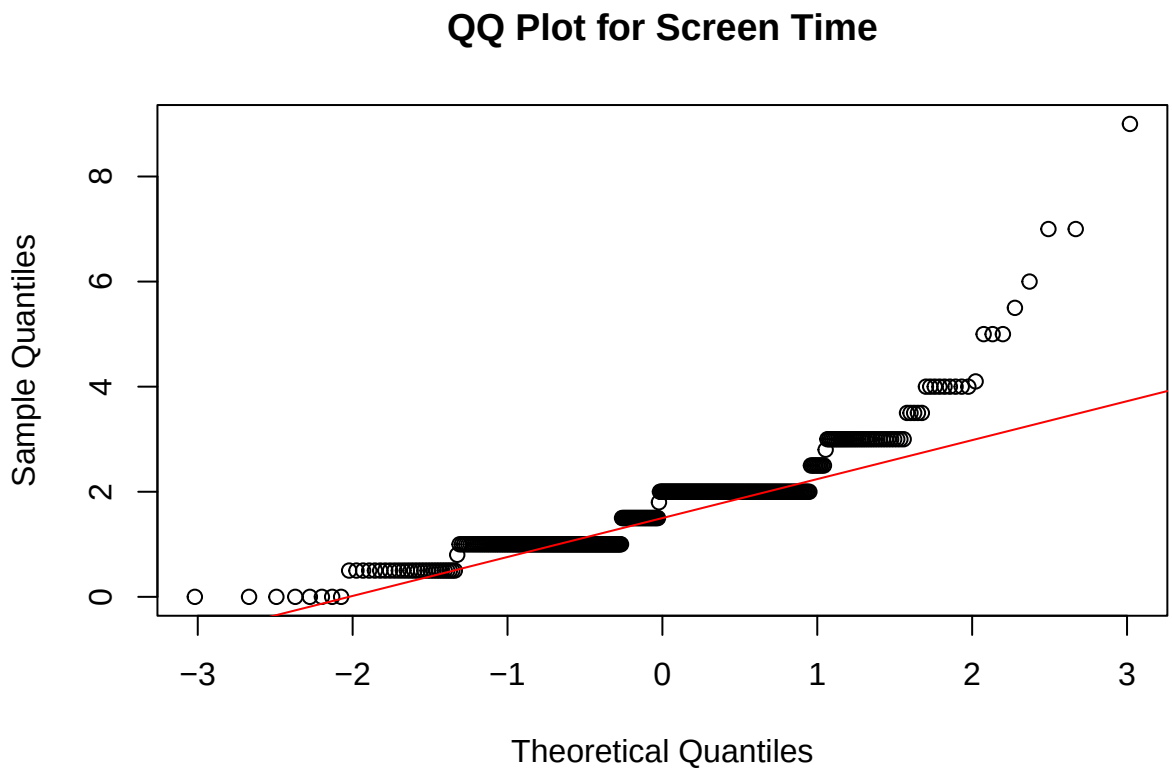
Independence:

The observations (children's daily screen times) must be independent of each other. This means that one child's screen time does not influence or is not influenced by another's.

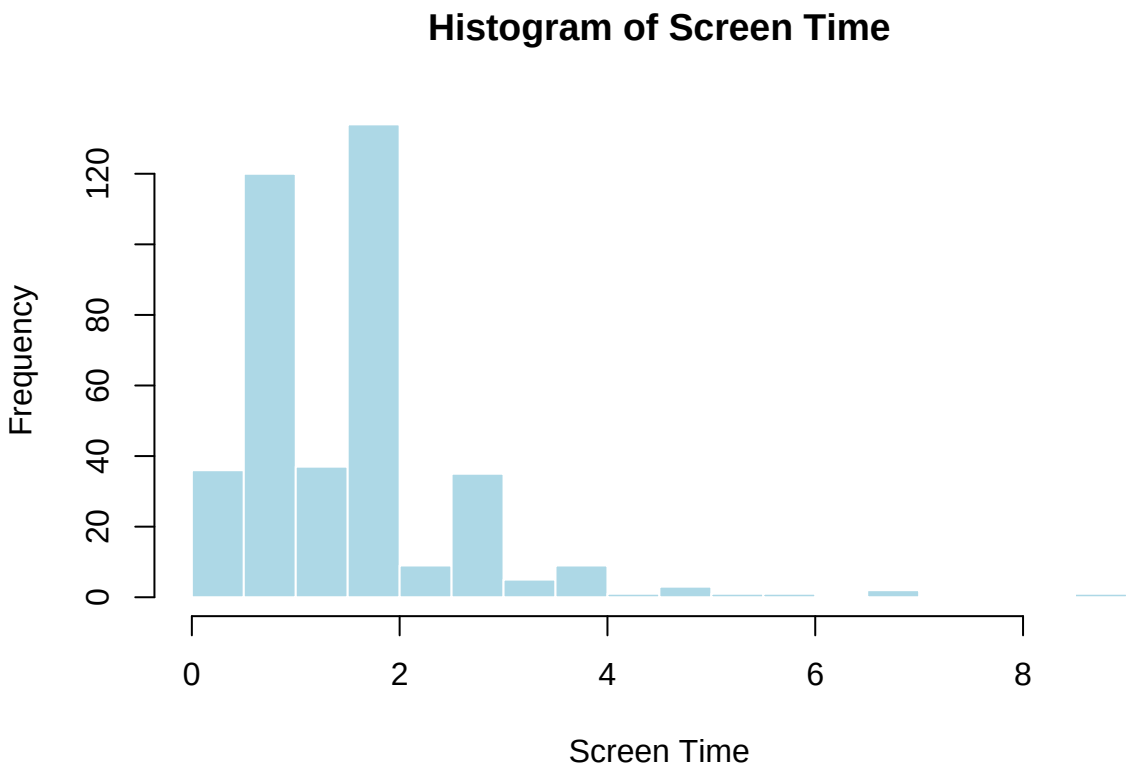
Normality:

The distribution of the sample mean should be approximately normal. This is especially crucial when the sample size is small. For larger samples (typically n > 30), the Central Limit Theorem assures us that the sampling distribution of the sample mean will be approximately normal regardless of the population distribution.

```
# QQ Plot for Screen Time
qqnorm(a3$screen, main = "QQ Plot for Screen Time")
qqline(a3$screen, col = "red")
```




```
# Histogram for Screen Time
hist(a3$screen, breaks = 30, main = "Histogram of Screen Time", xlab = "Screen Time", col = "lightblue"
```



Known or Large Sample Standard Deviation:

When the population standard deviation is unknown and the sample size is small, the distribution of the sample mean follows a t-distribution rather than a normal distribution.

Question 3

```
# Assuming 'screen' is the independent variable and 'physical' is the dependent variable
# Simple linear regression to explore the relationship between screen time and physical activity level
model <- lm(physical ~ screen, data = a3)

# Summary of the regression model
summary(model)
```

```
##
## Call:
## lm(formula = physical ~ screen, data = a3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.186 -3.380 -1.230  2.217 16.419
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.1862     0.4341  14.251   <2e-16 ***
## screen       -0.4033     0.2116  -1.905   0.0575 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.459 on 392 degrees of freedom
## Multiple R-squared:  0.009176,   Adjusted R-squared:  0.006648
## F-statistic:  3.63 on 1 and 392 DF,  p-value: 0.05747
```

```r
# Plotting to visualize the relationship
plot(a3$screen, a3$physical, main = "Physical Activity Level vs. Screen Time", xlab = "Screen Time", yla
abline(model, col = "red")
```