Story_Assignment_5

2024-11-13

```
data <- read.csv("bestroute.csv")
print(data)
```

```
##    HighwayMin HighwaySec BackRouteMin BackRouteSec
## 1          11         52           11           59
## 2           8         21            6           37
## 3           9         34            8           34
## 4          11         20           12           42
## 5          12         57           15            2
## 6          11         48           10           49
## 7           9         24            8            6
## 8           9         25            7           51
## 9           7         57           10           23
## 10          6         54            9            5
## 11         11         53           11           29
## 12          9         38            7           19
## 13         12         39            9           52
## 14          8         10            5            1
## 15          8         43            7           49
## 16          8         42            6            7
## 17         12          6            9           48
## 18         11         19            8           37
## 19          7         21            4           48
## 20          8         40            7           11
## 21         10         38           12           29
## 22          9         27            9           37
## 23         13         37           11           19
## 24          9         17           11            0
## 25         12          1            9           44
## 26         11         29            8           13
## 27          9         11            8            3
## 28         11         54           12           40
## 29          7         35            4           32
## 30          7         36            9            2
```

```
Highway_Min <- data$HighwayMin
Highway_Sec <- data$HighwaySec
BackRoute_Min <- data$BackRouteMin
BackRoute_Sec <- data$BackRouteSec


hwaytime <- Highway_Min * 60 + Highway_Sec
backtime <- BackRoute_Min * 60 + BackRoute_Sec
```

```r
time_difference <- hwaytime - backtime
print(time_difference)
```

```
##  [1]   -7  104   60  -82 -125   59   78   94 -146 -131   24  139  167  189   54
## [16]  155  138  162  153   89 -111  -10  138 -103  137  196   68  -46  183  -86
```

```r
mean_diff <- mean(time_difference)
sd_diff <- sd(time_difference)
n <- length(time_difference)

t_stat_long <- mean_diff / (sd_diff / sqrt(n))
print(t_stat_long)
```

```
## [1] 2.564499
```

```r
t_stat <- 2.5645
df = n-1
p_value <- 2 * pt(-abs(t_stat), df)
cat("p-value:", p_value, "\n")
```

```
## p-value: 0.01577428
```

```r
t_test_result <- t.test(hwaytime, backtime, paired = TRUE)
print(t_test_result)
```

```
## 
##  Paired t-test
## 
## data:  hwaytime and backtime
## t = 2.5645, df = 29, p-value = 0.01577
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  10.39416 92.27251
## sample estimates:
## mean difference 
##        51.33333
```

```r
Highway_Min <- data$HighwayMin
Highway_Sec <- data$HighwaySec
BackRoute_Min <- data$BackRouteMin
BackRoute_Sec <- data$BackRouteSec



hwaytime <- Highway_Min * 60 + Highway_Sec
backtime <- BackRoute_Min * 60 + BackRoute_Sec



time_difference <- hwaytime - backtime
print(time_difference)
```

```
##  [1]   -7  104   60  -82 -125   59   78   94 -146 -131   24  139  167  189   54
## [16]  155  138  162  153   89 -111  -10  138 -103  137  196   68  -46  183  -86
```

```r
mean_diff <- mean(time_difference)
sd_diff <- sd(time_difference)
n <- length(time_difference)

```

```r
ranked_diff <- rank(abs(time_difference))
signed_ranks <- ifelse(time_difference > 0, ranked_diff, -ranked_diff)

W_stat <- sum(signed_ranks[signed_ranks > 0])
#A
wilcox_test_1<- wilcox.test(hwaytime, backtime, paired = TRUE, exact = TRUE)
```

```
## Warning in wilcox.test.default(hwaytime, backtime, paired = TRUE, exact =
## TRUE): cannot compute exact p-value with ties
```

```r
print(W_stat)
```

```
## [1] 349
```

```r
print(wilcox_test_1)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  hwaytime and backtime
## V = 349, p-value = 0.01703
## alternative hypothesis: true location shift is not equal to 0
```

```r
#B
z_value <- (W_stat - (length(time_difference) * (length(time_difference) + 1) / 4)) /
           sqrt(length(time_difference) * (length(time_difference) + 1) * (2 * length(time_difference)
p_value_1 <- 2 * pnorm(-abs(z_value))  # Two-tailed test

print(W_stat)
```

```
## [1] 349
```

```r
print(wilcox_test_1)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  hwaytime and backtime
## V = 349, p-value = 0.01703
## alternative hypothesis: true location shift is not equal to 0
```

```r
print(p_value_1)
```

```
## [1] 0.01656553
```

```r
#C
wilcox_test_2 <- wilcox.test(hwaytime, backtime, paired = TRUE, exact=TRUE)
```

```
## Warning in wilcox.test.default(hwaytime, backtime, paired = TRUE, exact =
## TRUE): cannot compute exact p-value with ties
```

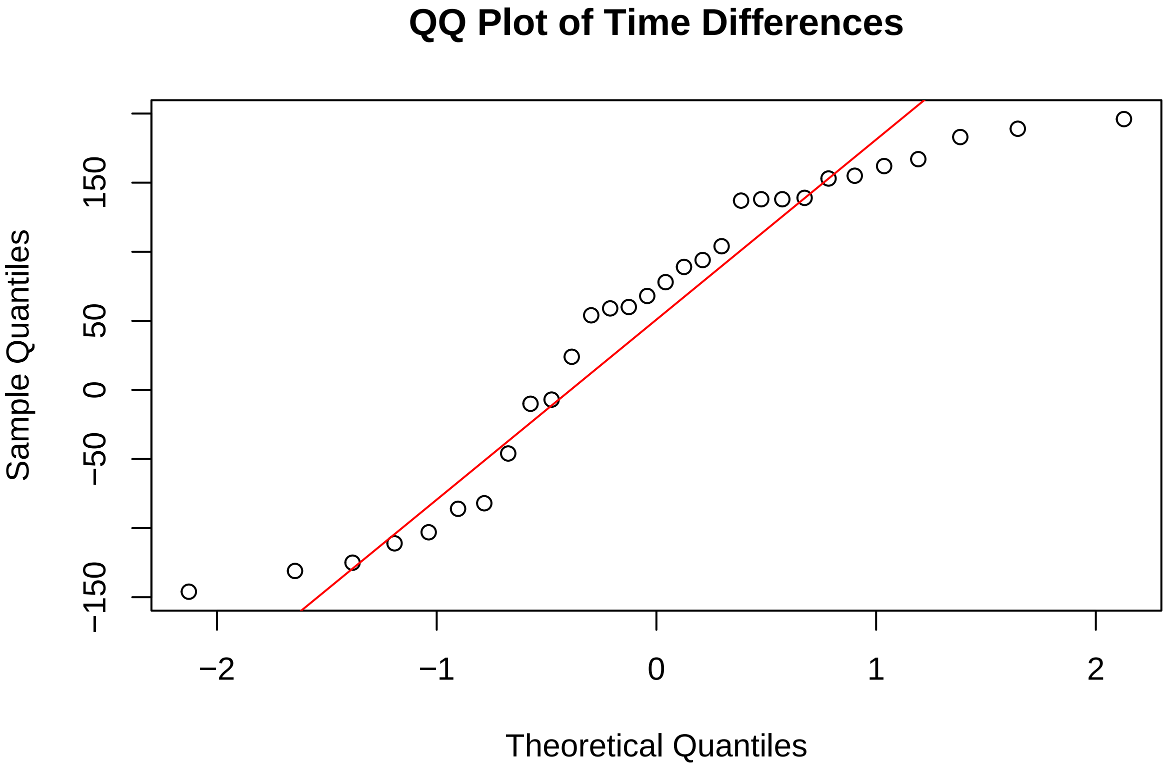
```r
print(wilcox_test_2)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  hwaytime and backtime
## V = 349, p-value = 0.01703
## alternative hypothesis: true location shift is not equal to 0
```

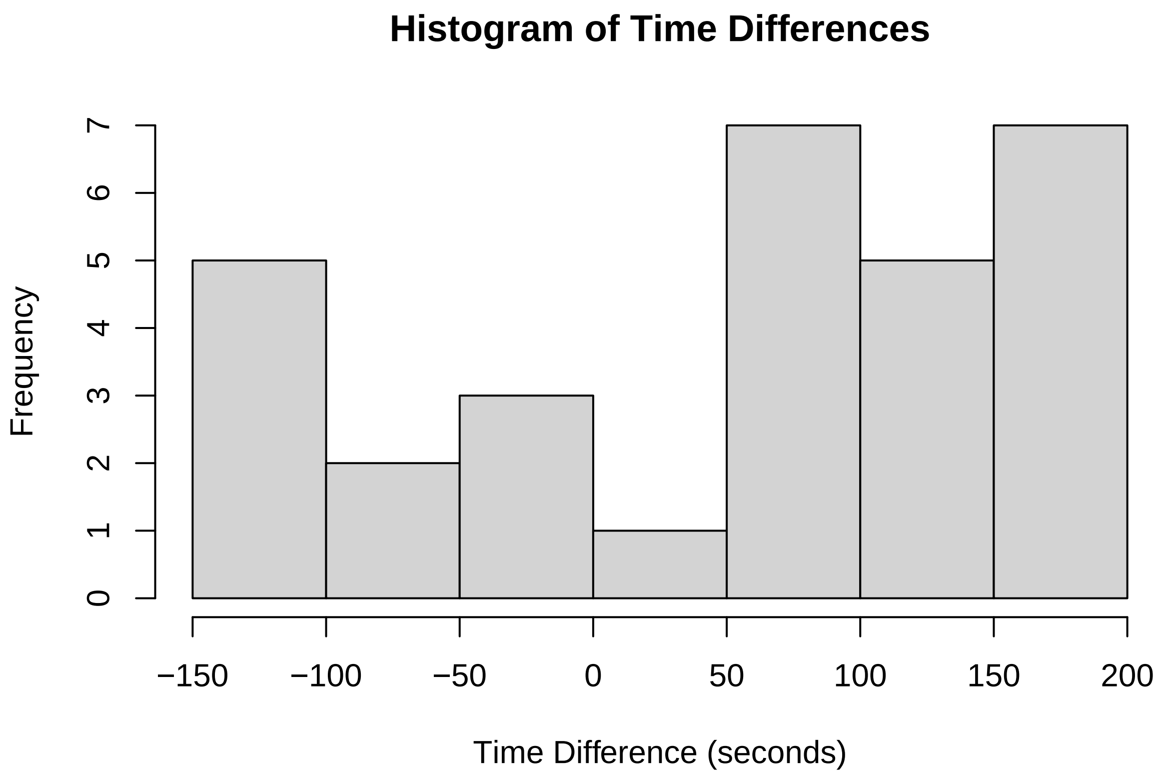
```
#The results from 2a, 2b, and 2c may slightly differ due to the method of approximation used. The norma
```

```
time_difference <- hwaytime - backtime

qqnorm(time_difference, main = "QQ Plot of Time Differences")
qqline(time_difference, col = "red")
```



```
hist(time_difference, main = "Histogram of Time Differences", xlab = "Time Difference (seconds)", breaks
```



```
shapiro_test <- shapiro.test(time_difference)

cat("Shapiro-Wilk Test p-value:", shapiro_test$p.value, "\n")
```

```
## Shapiro-Wilk Test p-value: 0.01348978
```

```
#3A: The data is not noramlly distributed becuase the QQ blot data does not follow a straight line, the

#3B: Your roomate was right. The correct test to determine this is the Wilcoxon signed-rank test and th
```

```
observed_counts <- c(Banana = 23, Vinegar = 18, Honey = 5, Sugar_Water = 10)

total_flies <- sum(observed_counts)

expected_counts <- rep(total_flies / length(observed_counts), length(observed_counts))
chi_square_stat <- sum((observed_counts - expected_counts)^2 / expected_counts)

df <- length(observed_counts) - 1

p_value <- pchisq(chi_square_stat, df, lower.tail = FALSE)

print(chi_square_stat)
```

```
## [1] 13.85714
```

```
print(p_value)
```

```
## [1] 0.003106198
```

```
chi_square_test <- chisq.test(observed_counts, p = rep(1/length(observed_counts), length(observed_count
print(chi_square_test)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  observed_counts
## X-squared = 13.857, df = 3, p-value = 0.003106
```

```
data <- matrix(c(16, 13, 5, 1), nrow = 2, byrow = TRUE)
colnames(data) <- c("Offspring_No", "Offspring_Yes")
rownames(data) <- c("Maternal_No", "Maternal_Yes")
data_1 <- as.table(data)

observed <- as.vector(data)

row_sums <- rowSums(data_1)
col_sums <- colSums(data_1)
total <- sum(data_1)

expected <- outer(row_sums, col_sums) / total

chi_square_stat <- sum((observed - expected)^2 / expected)
df <- (nrow(data_1) - 1) * (ncol(data_1) - 1)

p_value_no_correction <- pchisq(chi_square_stat, df, lower.tail = FALSE)

print(chi_square_stat)
```

```
## [1] 1.64272
```

```
print(p_value_no_correction)
```

```
## [1] 0.1999526
```

```
yates_correction <- sum((abs(observed - expected) - 0.5)^2 / expected)

p_value_yates <- pchisq(yates_correction, df, lower.tail = FALSE)
print(yates_correction)
```

```
## [1] 0.6788793
```

```
print(p_value_yates)
```

```
## [1] 0.4099729
```

```
chi_square_test_r <- chisq.test(data_1)
```

```
## Warning in chisq.test(data_1): Chi-squared approximation may be incorrect
```

```
cat("chi-square statistic:", chi_square_test_r$statistic, "\n")
```

```
## chi-square statistic: 0.6788793
```

```
cat("p-value:", chi_square_test_r$p.value, "\n")
```

```
## p-value: 0.4099729
```

```
cat("Yates continuity correction used by default in R:", chi_square_test_r$method, "\n")
```

```
## Yates continuity correction used by default in R: Pearson's Chi-squared test with Yates' continuity
```

```
#5d: There is no significant association between maternal anadromy and offspring anadromy becasue of th
```